UNIVERSITE LIBRE DE BRUXELLES

Faculté des Sciences Sociales, Politiques et Economiques

Année academique 2006-2007

Essays on the Dynamics of

Cross-Country Income Distribution

and Intra-Household Time Allocation

Thése présentée en vue de l'obtention du grade de

Docteure en Sciences Economiques par

Gisèle HITES

Sous la direction de Prof. Catherine DEHON

MEMBERS OF THE JURY

Thesis director:	Prof. Catherine Dehon, Université Libre de Bruxelles
Restricted Jury:	Prof. Natalie Chen, University of Warwick
	Prof. Paula Conconi, Université Libre de Bruxelles
	Prof. Alain Desdoigts, Université de Bourgogne & Université Paris I
	Prof. Davy Paindavaine, Université Libre de Bruxelles
	Prof. Robert Plasman, Université Libre de Bruxelles

ACKNOWLEDGEMENTS

I would like to begin my acknowledgements by thanking those that gave me that last big push that has propelled me to where I am right now, sitting in my office at the IESEG in Lille, finishing the final touches to the doctoral thesis that is to be submitted before Christmas.

The last big push started with Marina Emiris dragging me into the office on a Sunday early this year so that I could explain to her exactly what I had already done, what I still needed to do, and when I was going to do it. I do not want to imagine where I would be in my thesis had she not done so. Thank you, Marina.

Catherine Dehon picked up the relay, providing me with generous amounts of high-quality feedback on my work. She met with me whenever I needed to see her, for however long I needed to see her. She believed in me, providing me with positive encouragement throughout. Without her, I would never have known what it was to have a supervisor. Thank you, Catherine.

And then there was my family. As I took up the challenge of finishing the thesis without letting go of my many ‘extra’ jobs, I had to effectively let go of pretty much all of my family time. That family time was time spent taking care of my four (beautiful and intelligent) children, and time spent having fun with the family. The care was handed over to my husband, Gerald Ernst, who has had to redefine himself and to reorganize his life in order to take over charge. It was not easy. Many times it was hard. But he did it, and I am forever grateful. Thank you, Gerald. And the fun was put on hold, for me in any case. My children have all been most patient with me (“Mama, just go work, that way you will finish faster and then we will have time to play together”). I just hope that the number of pages is sufficient to justify all of my absences (“Mama, how long is your book going to be, how many pages?”). Thank you, Elkyn, Ilaria, Ryoko and Kjetil. This thesis is dedicated to you four monsters.

I would like to continue my acknowledgement by thanking those that have helped me along the lengthy road preceding the last big push.

At the top of that list, I have to thank my parents, Leslie and László Hites, for always being there to help me, in whatever way I needed it. I could count on them for basically anything: moral support, financial support, and even childcare despite their California residence and overloaded work schedules. I would also like to thank them for instilling in their children the ethic of hard work, because if I am where I am today, it is because I worked hard to get there. And I learned how to work hard by watching them work hard. I am extraordinarily lucky to have parents like them. Thank you, Mami and Papi.

Then I have to also thank all five of my siblings, Maya, Romina, Nathalie, Renée and Damian Hites, for the part that they have each played at one point or another in helping me out in my everyday life as a poor doctoral student in a commuting relationship with young children. Sympathetic ears were lent hearing me out on the travails of the thesis, and time was spent taking care of my children. Thank you, sisses and bro.

And I do not want to forget those with whom I shared my days at ECARES. I am talking about my officemates, Natalie Chen, Marina Emiris, Joelle Liebermann, Antonello d'Agostino, and Paulo Casini, who all contributed to making life in the office a more pleasant and less lonely affair. Thank you, guys.

Last, but not least, I would like to thank those of you at ULB who have believed in me and stuck up for me over the years. Having chosen a rather unconventional path, I did not emit the standard signals that inspire confidence in an academic setting. Some people were able to see through the noise, recognize the researcher underneath, and stick their necks out to get my re-inscription forms signed when they needed signing, contracts renewed when they needed renewing, etc. Without such support, I would not have been able to complete my thesis. For this, I would like to thank both Catherine Dehon and Marjorie Gassner, knowing very well that other people also contributed to this behind-the-scenes support that is unfortunately very often invisible to the beneficiary. Thank you to those of you who also contributed to this support, and are not mentioned.

ABSTRACT OF THE THESIS

This thesis contributes to two completely unrelated debates in the economic literature, similar only in the relatively high degree of controversy characterizing each one.

The first part is methodological and macroeconomic in nature, addressing the question of whether the distribution of income across countries is converging (i.e. are the poor catching up to the rich?) or diverging (i.e. are we witnessing the formation of two exclusive clubs, one for poor countries and another one for rich countries?). Applications of the simple Markov model to this question have generated evidence in favor of the divergence hypothesis. In the first chapter, I critically review these results. I use statistical inference to show that the divergence results are not statistically robust, and I explain that this instability of the results comes from the application of a model for discrete data to data that is actually continuous. In the second chapter, I reposition the whole convergence-divergence debate by placing it in the context of Silverman's classic survey of non-parametric density estimation techniques. This allows me to use the basic notions of fuzzy logic to adapt the simple Markov chain model to continuous data. When I apply the newly adapted Markov chain model to the cross-country distribution question, I find evidence against the divergence hypothesis, and this evidence is statistically robust.

The second part of the thesis is empirical and microeconomic in nature. I question whether observed differences between husbands' and wives' participation in labor markets are due to different preferences or to different constraints. My identification strategy is based on the idea that the more power an individual has relative to his/her partner, the more his/her actions will reflect his/her preferences. I use 2001 PSID data on cohabiting couples to estimate a simultaneous equations model of the spousal time allocation decision. My results confirm the stylized fact that specialization and trade does not explain time allocation for couples in which the wife is the primary breadwinner, and suggest that power could provide a more general explanation of the observations. My results show that wives with relatively more power choose to work more on the labor market and less at home, whereas husbands with more power choose to do the opposite. Since women start out from a lower level of labor market participation than men do, it would seem that spouses' agree that the ideal mix of market work and housework lies somewhere between the husbands' and the wives' current positions.

TABLE OF CONTENTS

GENERAL INTRODUCTION

This thesis is composed of two very different parts that share one common element. The first part is methodological and macroeconomic and the second part is empirical and microeconomic, but both parts address very basic questions that have been the source of much controversy in the economic literature. In the first part, the question is whether or not the distribution of income across countries is converging, and in the second part, the question is whether or not labor market outcomes reflect labor market preferences for women.

Part I: The dynamics of cross-country income distribution

The simple Markov chain model has been widely used in the social sciences to study the phenomenon of mobility. Empirical applications have included geographic, labor, and social mobility. The prime attraction in this approach lies in the simplicity with which short-run dynamics are translated into long-run tendencies. One active area of research in which the short and long-run properties of a panel of data are of particular interest is the study of the distribution of income across countries. In this body of literature, the simple Markov chain model plays a particularly controversial role. Quah uses this model to paint a new picture of the world in which the rich and the poor are diverging to form 'twin peaks'.

Chapter1: On the robustness of the twin-peaked ergodic distribution of income across countries

In the first chapter, I contest the robustness of the twin peaks conclusion, questioning the suitability of this discrete model to analyse continuous data on income. First, I use a filter to clean the data of any short-run noise. This procedure enhances the twin-peaked result. Second, I introduce the use of statistical inference to the debate and show that the long-run distribution is exceedingly unstable. The confidence region for the estimated twin-peaked long-run distribution includes unimodal and trimodal distributions. Third, I derive an analytical expression defining this instability. This expression reveals that the shape of the long-run distribution is entirely determined by the relatively few observations of country mobility, and not at all determined by the relatively many observations of country immobility. So, I take a closer look at the few observations of mobility, and realize that many of these observations are not really observations of mobility at all, but rather observations of short-run noise. This short-run noise is generated by the discretization of the continuous income variable that is necessary in order to apply the discrete Markov chain model. The filter applied in the first section cleans up this short-run noise, but this exacerbates the fragility of the ergodic distribution. In Chapter 2, fuzzification provides a solution to this problem.

Chapter 2: Fuzzifying the cross-country income convergence debate

In the second chapter, I study the causes underlying this breakdown of the simple Markov chain model and I propose a simple solution that uses notions imported from the domain of fuzzy logic to adapt the simple Markov chain model to continuous data.

I begin by redefining the cross-country income convergence debate in two manners. First, I return to first principles, reviewing the basics of Markov chain modelling. The literature has concentrated its efforts on space-discretization, completely ignoring the issue of time-discretization. However, discrete observations of a fundamentally continuous process can only be modelled as a Markov chain if space and time discretization has been carried out properly. Second, I place the issue of space-discretization in the context of Silverman's classic survey of non-parametric density estimation techniques. This view of the transition matrix estimation process reveals some very basic drawbacks in the most common estimation techniques and suggests easy ways of improving upon these techniques. The fuzzification of the simple Markov chain model that is developed in the rest of the chapter represents one such improvement.

Taking this non-parametric view of the estimation process, I examine the short-run noise generated by the application of a discrete model to continuous data. This short-run noise can be classified into two groups, the first one containing observations of short-run dynamics and the second one containing observations tainted by measurement error. In order to address these two very different problems, fuzzification of the simple Markov chain model is carried out in two different ways. First, instead of using point income levels to define the different income classes, intervals of income are used to delineate the frontiers between the different classes of income. The income class frontiers are fuzzified. Second, point observations of income are replaced by distributions across intervals of income, the widths of which are proportional to the quality of the underlying data. The income observations are fuzzified.

When fuzzification is carried out (and other corrections are made), the long-run distribution of income across countries becomes unimodal, and this time the results are statistically robust.

Part II: The dynamics of intra-household time allocation

Chapter 3: Nature versus Nurture: An empirical analysis of the division of labor within American couples

Less women than men participate in the American labor market, and when women do participate, they work less hours on average than men. Why are women less active on average than men on the labor market? Is it nature or nurture, choice or constraint? My results show that observed labor market participation does not reflect true preferences.

I use data from the Panel Study of Income Dynamics (PSID) on cohabiting couples for 2001 and show that the effective labor market supply of the women is constrained at levels that are lower than desired. In order to identify true preferences I assume that the more power an individual has relative to his/her partner, the more his/her actions will reflect his/her preferences. Four proxies for power are used: age, education, wage and local sex ratios. A system of four simultaneous equations is used to model the couples' allocation of time between house and market work. The regressions generate the following results.

For American 'wives,' an increased share of the total education and wages for the couple is accompanied by an increased probability of working *longer* hours on the labor market and *shorter* hours at home. For example, a 'typical' woman with 4 years of college education and an average wage, who is living with a man with less education and a lower wage, has 14% less chance of working part-time and 11% more chance of working full-time when compared to the same woman who is living with a man with equal education and a similar wage. So this same woman will most probably work full-time when she is in the relatively more powerful situation with respect to her partner, and part-time when she is in the less powerful situation. Note that this is not a specialization and trade story because for this to be the case, the opposite would have to hold true for men, and this is not the case.

For American 'husbands,' an increased concentration of women in the population is accompanied by an increased probability of working *shorter* hours on the labor market and *longer* hours at home. For example, a 'typical' man who is living in a state where 50% of the population in his age-bracket is female has 10% less chance of working 40-45 hours a week and 10% more chance of working more than 45 hours a week, when compared to the same man living in a state where 51% of the population in his age-bracket is female. So this same man will most probably have a 'small' full-time job when he is in the more powerful situation with respect to his partner and a 'big' full-time job when he is in the less powerful situation.

These results suggest that men and women are more similar in their preferences on the work-family trade-off than the labor market statistics reveal. Women do less market work and more housework than husbands, but they would like to do more market work and less housework. Men do more market work and less housework than women, but they would like to do less market work and more housework. Men and women seem to agree that the ideal place to be is somewhere in between.

PART I:

The Dynamics of Cross-Country Income Distribution

Chapter 1

On the Robustness of the Twin-peaked Ergodic Distribution of Income across Countries

Abstract: The very basics of statistical inference are surprisingly absent from empirical applications of the simple Markov chain model. This paper develops three tools in an effort to remedy the situation. Confidence intervals for the estimated transition probabilities are analytically derived, and confidence intervals for the resulting ergodic probabilities are numerically calculated. The notion of the elasticity of the ergodic probabilities with respect to the transition probabilities is developed and quantified. These tools are applied within the context of the cross-country income convergence debate. In the literature on convergence, the simple Markov chain model indicates evolution towards a twin-peaked world. Although cleansing the ergodic distribution of income across countries of short-run noise reinforces its twin-peaked shape, these twin peaks are not statistically significant. Moreover, the specific type high immobility reflected by the data on income renders the estimated transition matrix particularly prone to the generation of twin-peaked ergodic distributions.

Keywords: Ergodic distribution; Filters; Income distribution; Markov chains; Twin peaks.

JEL classification: C23; O57.

1. Introduction

The simple Markov chain model has been widely used in the social sciences to study the phenomenon of mobility. Empirical applications have included geographic, labor, and social mobility. The prime attraction in this approach lies in the simplicity of the characterization of the steady state. One active area of research in which the long run properties of a panel of data are of particular interest is the study of the distribution of income across countries. In this body of literature, Quah’s (1993a,b) application of the simple Markov chain model reveals evidence of a world in which the rich and the poor are diverging to form ‘twin peaks’. In what follows, the robustness of this conclusion is assessed.

The tools available for the assessment of the robustness of the twin-peaked shape of the ergodic distribution are surprisingly scarce. Statistical inference on Markov chains is basically limited to hypothesis testing on the transition probabilities. In this paper, I derive confidence intervals for the transition probabilities analytically, and confidence intervals for the ergodic probabilities numerically. When I calculate these confidence intervals for the cross-country income distribution application, I find that the confidence region for the transition probability matrix is small, but that the confidence region for the ergodic distribution is large enough to contain all sorts of different shaped distributions. I conclude that the twin-peaked result is not statistically robust.

This result highlights the fragility of the simple Markov chain model when applied in certain empirical settings. In this paper, I use the economic concept of elasticity to derive an analytical expression for this fragility. This expression allows me to quantify the fragility and to identify the causes of the fragility. I show that the certain type of high immobility displayed by the underlying data first imposes a very particular functional form upon the ergodic distribution, and then situates the analysis in the degenerate part of the domain of this particular ergodic function. Calculation of this elasticity could help researchers assess whether or not the simple Markov chain model is indeed appropriate for the modelling of their data.

In Section 1, the data is filtered of business cycle type fluctuations. In theory, the long-run distribution is independent of short-run noise; however, in practice, this is not the case because the sample is finite. The simple Markov chain model calls for discrete data, yet empirical applications are based on continuous data. This introduces short-run noise into the estimation process which in turn contaminates the long-run distribution (because the sample is finite). It is found that filtering the data reinforces the twin peaks result. In Section 2, statistical inference is carried out. Section 2.1 presents the analytical derivation of the confidence intervals for the transition probabilities. Section 2.2 presents the numerical calculations of the confidence intervals for the ergodic probabilities. In Section 3, the analytical expressions for the elasticities of the ergodic probabilities with respect to the transition probabilities are derived. A discussion of the implications of these expressions is carried out. In Section 4, conclusions are presented.

2. Filter

First, let us establish notation. There are a finite number of states m $(i = 1,...,m)$ and transitions between these states are observed at regular intervals for a finite length of time T $(t = 1,...,T)$. Let $N(t)$ be the matrix of observed transitions at time *t* where the *ij*th element is $n_{ij}(t)$ (the number of transitions from state *i* to state *j* observed at time *t*), and let $n(t)$ be the distribution of the observations across the states at time *t* where the *i*th element is $n_i(t)$ (the number of observations in state *i* at time *t*). We assume that the observed transitions are generated by a simple (i.e. of order one) time-homogenous Markov chain according to the matrix of transition probabilities *P* where the *ij*th element is p_{ij} (the probability of transiting from state *i* to state *j*). The maximum likelihood estimator of *P* is denoted $\hat{P}$, where the *ij*th element is:

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{j=1}^{m}\sum_{t=1}^{T} n_{ij}(t)}. \qquad (1)$$

The model can then be summarized by the following expression:

$$n(t+1) = n(t)P = n(0)P^{t+1}. \qquad (2)$$

In this paper, we are interested in the long-run tendencies of the distribution of the observations (i.e. the ergodic distribution), so we let t go to ∞, and we get:

$$n(\infty) = n(\infty)P = n(0)P^{\infty}. \tag{3}$$

The middle term tells us that the ergodic distribution is nothing other than the left eigenvector corresponding to the unit eigenvalue of the transition matrix. The term on the right tells us that the rows of the transition matrix converge to the ergodic distribution as t approaches ∞. These two pieces of information will be useful to us in Section 3.2.

Second, let us review Quah's application of the simple Markov chain model. The data used is the Laspeyres index of annual real per capita income from the Summers and Heston (1991) Penn World Tables for 118 countries (relative to the world average) for the 1962-84 time period. Five possible states are defined by discretizing the set of possible values of relative incomes into intervals at 1/4, 1/2, 1, and 2. The results are presented below:

$$\hat{P}_{Quah} = \begin{matrix} 0.97 & 0.03 & 0 & 0 & 0 \\ 0.05 & 0.92 & 0.04 & 0 & 0 \\ 0 & 0.04 & 0.92 & 0.04 & 0 \\ 0 & 0 & 0.04 & 0.94 & 0.02 \\ 0 & 0 & 0 & 0.01 & 0.99 \end{matrix} \tag{4}$$

$$n_{Quah}(\infty) = \begin{matrix} 0.24 & 0.18 & 0.16 & 0.16 & 0.27 \end{matrix}$$

The estimated ergodic distribution indicates an evolution towards a bipolar world of haves and have-nots.

Third, let us start out our analysis by reproducing Quah's results. Reestimation of the transition probability matrix and recalculation of the corresponding ergodic distribution using data from a more recent version of the Penn World Tables (i.e. Mark 5.6) turns out to be surprisingly revealing. The sample is composed of the 111 countries for which there is continuously available data for the period 1960-89, that is the 118 countries used by Quah (1993a,b) minus Afghanistan, Sudan, Ethiopia, Liberia, Nepal, Iraq, and Tanzania. The results are presented below:

$$\hat{P}_{Hites-0} = \begin{matrix} 0.96 & 0.04 & 0 & 0 & 0 \\ 0.06 & 0.91 & 0.03 & 0 & 0 \\ 0 & 0.03 & 0.95 & 0.02 & 0 \\ 0 & 0 & 0.03 & 0.95 & 0.02 \\ 0 & 0 & 0 & 0.01 & 0.99 \end{matrix} \tag{5}$$

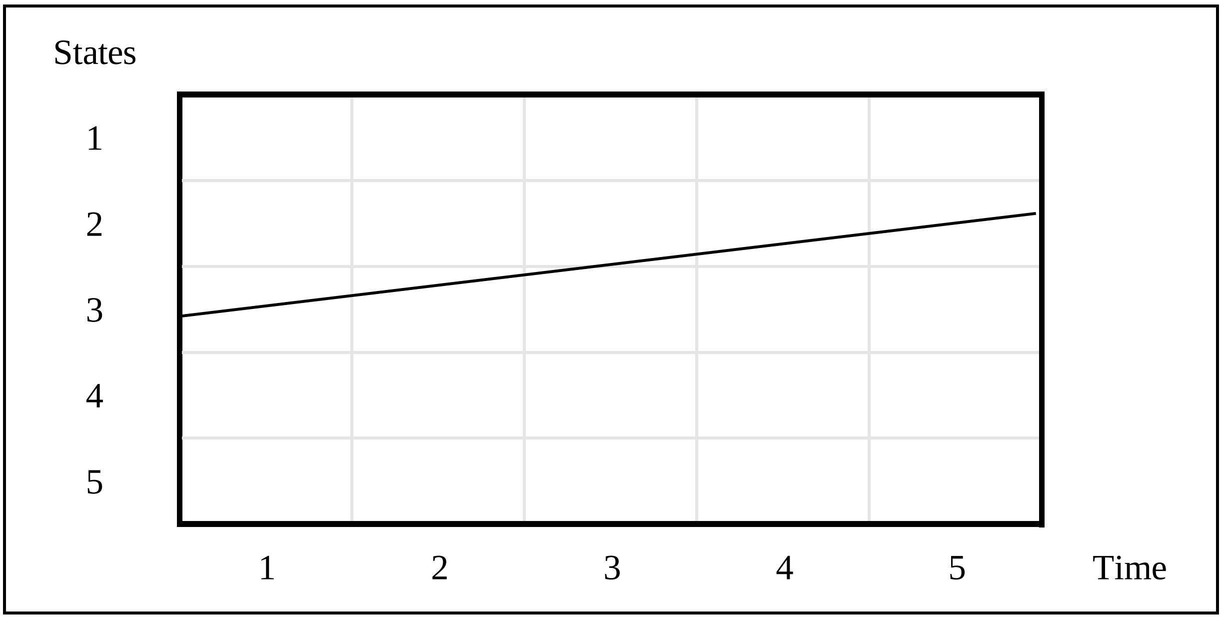
$$n_{Hites-0}(\infty) = 0.30 \quad 0.19 \quad 0.17 \quad 0.12 \quad 0.22$$

This estimated transition matrix is very similar to the one calculated by Quah, but it generates an ergodic distribution with a rather different economic interpretation. Although the ergodic distribution is still twin peaked, it is the poorest class, and not the richest class as before, that absorbs the biggest portion of the population. In other words, just data inaccuracies would constitute a sufficiently big perturbation to the estimated transition matrix to alter the ergodic distribution in an economically significant manner.

Finally, let us turn to the issue of filtering. In the simple Markov chain model, the estimated transition probability matrix is used to extract information concerning the mobility of countries within the distribution of incomes. This information is camouflaged by two sources of noise. The first is generated by inaccuracies in the data and will not be discussed here. The second results from using continuous data to estimate a discrete model (i.e. from translating continuous data into discrete data by defining income class frontiers) and will be discussed below. Whereas the first source of noise affects the inference in ways that are unknown to us, the second source of noise can be observed directly.

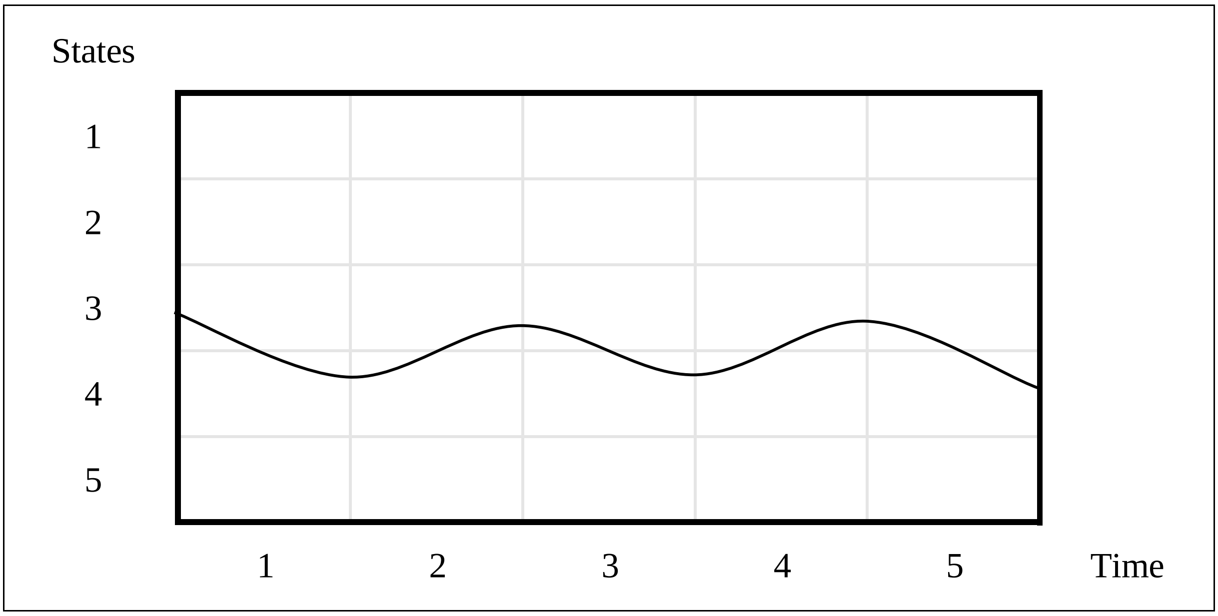
In the simple Markov chain model, transitions represent mobility. In reality, however, transitions can occur for two reasons. Transitions can result from higher (or lower) than average world growth in a country (see Figure 1).

Figure 1: Transitions resulting from higher than average world growth



This is what we call mobility and this is what we would like to measure. Transitions can also result from business cycle type variations in a country's income when the level of income is situated very close to one defining a frontier between classes (see Figure 2).

Figure 2: Transitions resulting from business cycle type variations



This is clearly not mobility and since such transitions are included in our calculation of mobility, it is necessary to purge the data of such noise. In sum, we need to correct for the bias that short run fluctuations in income introduce into the calculation of long run tendencies in the distribution of world incomes.[1]

Two remarks are in order. First, the noise introduced by business cycle type fluctuations at frontiers between income classes only affects estimates of the probability of transitions between adjacent states, that is estimates of the elements on the diagonals just above and below the main diagonal of the transition probability matrix. Given the tri-diagonal structure and the stochastic nature (i.e. all rows sum to one) of the estimated transition probability matrices, all elements of the estimated transition probability matrices are potentially affected by this noise. Second, although the business cycle type fluctuations at frontiers between income classes that do not constitute mobility but are recorded as such compensate for real mobility that occurs within classes and is therefore not recorded as such, the former over compensates for the latter (see Figures 3 to 6). We conclude from these two remarks that we would expect the elements on the main diagonal of the estimated transition probability matrix to be underestimated, the rest of the elements to be overestimated, and therefore mobility to be overestimated.

In order to cleanse the data, it is necessary to tighten the conditions under which a transition is considered to represent mobility. Here this is achieved by requiring a transition to last a minimum number of periods in order for it to be counted as mobility, this minimum number of periods being defined as just over the average number of periods spanned by a business cycle. Four increasingly fine filters are applied to the original data. The first filter counts transitions as mobility if they last for at least one year, that is if no transition is observed during the year following the initial transition. The second, third and fourth filters do the same for two, three and four year spans, respectively. The results obtained from application of these four filters are presented below:

[1] This is an estimation problem arising from the bias introduced into the transition matrix by fitting a discrete model to a finite sample of continuous data. In theory, the ergodic distribution is independent of short run noise.

$$\begin{aligned}
n_{Hites-0}(\infty) &= 0.30 \quad 0.19 \quad 0.17 \quad 0.12 \quad 0.22 \\
n_{Hites-1}(\infty) &= 0.42 \quad 0.17 \quad 0.14 \quad 0.08 \quad 0.19 \\
n_{Hites-2}(\infty) &= 0.41 \quad 0.12 \quad 0.12 \quad 0.09 \quad 0.26 \\
n_{Hites-3}(\infty) &= 0.52 \quad 0.13 \quad 0.10 \quad 0.07 \quad 0.19 \\
n_{Hites-4}(\infty) &= 0.51 \quad 0.14 \quad 0.09 \quad 0.07 \quad 0.20
\end{aligned} \tag{6}$$

Two observations are interesting to note. First, of the 127 transitions initially counted, almost half are due to short-term fluctuations in income and not the long-term tendencies that we are trying to measure. Indeed, of the 127 transitions initially counted, 29 (15, 12 and 4, respectively) lasted less than two (three, four and five, respectively) years (see Figure 7). Second, the application of increasingly fine filters reinforces the twin peaked shape of the ergodic distribution. In very rough terms, comparing the ergodic distribution calculated from the most finely filtered data to the ergodic distribution calculated from the unfiltered data, the part of the distribution falling into the poorest class (left hand peak) increases by 20% to reach 50%, the part of the distribution falling into the richest class (right hand peak) remains at 20%, and the part of the distribution falling into the middle classes decreases uniformly.

In sum, when the data is purged of short run noise, the evidence for twin peaks is reinforced. But just how robust is this evidence? In the next section, statistical inference is carried out.

3. Statistical inference

The results presented in the previous section are not complete. What is missing is information on the precision of these results. In most empirical contexts, point estimates are necessarily reported along with their confidence intervals. In this particular context of applied Markov chain analysis, such information is intriguingly never provided (see any of the references cited in the bibliography), a point made by Reichlin (1999) in her discussion of Quah (1999). In this section, the confidence region for the ergodic distribution is derived. This is done in two steps. In Section 3.1, the confidence intervals for each of the transition probabilities are first derived. In Section 3.2, these confidence intervals are then mapped onto the space inhabited by the ergodic distribution. Because of the multidimensionality of the problem, this mapping is carried out numerically, rather than analytically.

3.1 Transition probability matrix

In this paper I am concerned about the robustness of the twin-peaked shape of the ergodic distribution. Variations in the ergodic distribution are generated by variations in the elements of the transition probability matrix. Therefore, the confidence region for the ergodic distribution depends upon the confidence region for the transition probability matrix. Anderson and Goodman (1957) study the asymptotic properties of first-order Markov chains and derive such a confidence region. They show that, for each state i, under the null hypothesis $\hat{p}_{ij} = \tilde{p}_{ij} \quad (i, j = 1,..., m)$:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} n_i^T \frac{(\hat{p}_{ij} - \tilde{p}_{ij})^2}{\tilde{p}_{ij}} \sim c^2(m(m-1)) \ , \tag{7}$$

where $n_i^T = \sum_{t=0}^{T} n_i(t)$ and $\tilde{p}_{ij}$ are the theoretical transition probabilities under the null hypothesis. This allows them to define the confidence region as the set of $\tilde{p}_{ij}$ for which Equation (7) is less than or equal to the a significance point of the relevant chi-square distribution:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} n_i^T \frac{(\hat{p}_{ij} - \tilde{p}_{ij})^2}{\tilde{p}_{ij}} \leq c_{1-a}^2 (m(m-1)) \ . \tag{8}$$

In what follows, we will use these results to explicitly define the limits of the confidence intervals for the individual transition probabilities p_{ij}.

These results need to be slightly adapted to the situation at hand. Equation (7) is derived under the assumption that every $p_{ij} > 0$, which is clearly not the case here.[2] Nevertheless, we can circumvent this problem by simply neglecting the zeros in each row. Indeed, we are not interested in testing whether these elements are truly zero (i.e. I accept that certain degrees of income mobility do not occur.); rather, we are interested in testing whether

[2] Anderson and Goodman's test statistic is an asymptotic result, valid only for sufficiently large samples (i.e. those for which the expected frequency in each of the cells of the transition matrix is superior to 5).

the positive elements are significantly different from some theoretically interesting values (i.e. I want to evaluate whether the observations of mobility are compatible with certain stories.). So, if I discard the three zeros present in the first row of the estimated transition probability matrix, I am left with two values that sum to one. The remains of this first row look like a row of an estimated transition probability matrix resulting from the estimation of a two-state Markov chain model, and I treat it as such. I do the same for the remaining rows. The first and the fifth rows of the estimated transition matrix are treated as the results from the estimation of a two-state Markov chain model (i.e. $m_1 = m_5 = 2$), and the second to fourth rows are treated as the results from the estimation of a three-state Markov chain (i.e. $m_2 = m_3 = m_4 = 3$). Each row has a c^2 distribution with $m_i - 1$ degrees of freedom and since these rows are asymptotically independent, they can be added to obtain a c^2 distribution with $\sum_{i=1}^{m}(m_i - 1)$ degrees of freedom. So, *m* varies from row to row, and Equation (7) becomes:

$$\sum_{i=1}^{m}\sum_{j=1}^{m_i} n_i^T \frac{(\hat{p}_{ij} - \tilde{p}_{ij})^2}{\tilde{p}_{ij}} \sim c^2 \Big(\sum_{i=1}^{m}(m_i - 1)\Big) \ . \tag{9}$$

Replacing the theoretical $\tilde{p}_{ij}$ in the denominator of the test statistic in Equation (9) by the empirical $\hat{p}_{ij}$ as suggested in Anderson and Goodman (1957), and marking with an * the coordinates of the transition probability for which we would like to construct the confidence interval, Equation (8) becomes:

$$n_{i^*}^T \frac{(\hat{p}_{i^*j^*} - \tilde{p}_{i^*j^*})^2}{\hat{p}_{i^*j^*}} + \sum_{\substack{i=1 \\ (i,j)\neq}}^{m} \sum_{\substack{j=1 \\ (i^*,j^*)}}^{m_i} n_i^T \frac{(\hat{p}_{ij} - \tilde{p}_{ij})^2}{\hat{p}_{ij}} \le c_{1-a}^2 \Big(\sum_{i=1}^{m}(m_i - 1)\Big) \ . \tag{10}$$

Rearranging terms in the expression in order to isolate $\tilde{p}_{i^*j^*}$ gives us:

$$\tilde{p}_{i^*j^*}^{\inf} \le \tilde{p}_{i^*j^*} \le \tilde{p}_{i^*j^*}^{\sup} , \tag{11}$$

where:

$$\tilde{p}_{i^*j^*}^{\inf} = \hat{p}_{i^*j^*} - \sqrt{\frac{\hat{p}_{i^*j^*}}{n_{i^*}^T}\left(c_{1-a}^2 \left(\sum_{i=1}^{m}(m_i - 1)\right) - \sum_{\substack{i=1 \\ (i,j)\neq}}^{m} \sum_{\substack{j=1 \\ (i^*,j^*)}}^{m_i} n_i^T \frac{(\hat{p}_{ij} - \tilde{p}_{ij})^2}{\hat{p}_{ij}} \right)}$$

$$\tilde{p}_{i^*j^*}^{\sup} = \hat{p}_{i^*j^*} + \sqrt{\frac{\hat{p}_{i^*j^*}}{n_{i^*}^T}\left(c_{1-a}^2 \left(\sum_{i=1}^{m}(m_i - 1)\right) - \sum_{\substack{i=1 \\ (i,j)\neq}}^{m} \sum_{\substack{j=1 \\ (i^*,j^*)}}^{m_i} n_i^T \frac{(\hat{p}_{ij} - \tilde{p}_{ij})^2}{\hat{p}_{ij}} \right)}.$$

Notice that this expression does not actually explicitly define the limits of the confidence interval for the unknown true value of the transition probability in question, $\tilde{p}_{i^*j^*}$. These limits are defined in terms of the unknown true values of the other transition probabilities, $\tilde{p}_{ij}$. And so we are faced with a system of $\sum_{i=1}^{m}(m_i - 1)$ simultaneous equations. This analytical complexity is due to the multidimensionality of the problem. In trying to derive confidence intervals for the individual transition probabilities, we are trying to map out a multidimensional confidence region onto its component dimensions.

The analytical complexity of this problem can be overcome by decomposing Equation (11) into its individual dimensions. Because the rows of a Markovian transition matrix are independent, we can consider the individual rows of the transition matrix separately, and Equation (10) reduces to:

$$n_{i^*}^T \frac{(\hat{p}_{i^*j^*} - \tilde{p}_{i^*j^*})^2}{\hat{p}_{i^*j^*}} + \sum_{j=1}^{m_{i^*}} n_{i^*}^T \frac{(\hat{p}_{i^*j} - \tilde{p}_{i^*j})^2}{\hat{p}_{i^*j}} \leq c_{1-a}^2 (m_{i^*} - 1) \quad . \tag{12}$$

Because our interest lies with the generic individual transition probability $p_{i^*j^*}$, we can lump together the other transition probabilities in row i^* into the term $1 - p_{i^*j^*}$, and Equation (12) reduces to:

$$n_{i^*}^{T}\frac{(\hat{p}_{i^*j^*}-\tilde{p}_{i^*j^*})^2}{\hat{p}_{i^*j^*}}+n_{i^*}^{T}\frac{((1-\hat{p}_{i^*j^*})-(1-\tilde{p}_{i^*j^*}))^2}{(1-\hat{p}_{i^*j^*})}\leq c_{1-a}^{2}(2-1)\,. \tag{13}$$

Setting $a = 0.05$, rearranging terms and simplifying, we get:

$$\hat{p}_{i^*j^*}-1.96\sqrt{\frac{\hat{p}_{i^*j^*}(1-\hat{p}_{i^*j^*})}{n_{i^*}^{T}}}\leq\tilde{p}_{i^*j^*}\leq\hat{p}_{i^*j^*}+1.96\sqrt{\frac{\hat{p}_{i^*j^*}(1-\hat{p}_{i^*j^*})}{n_{i^*}^{T}}}\,. \tag{14}$$

Applying Equation (14) to the estimated transition probability matrix in Equation (5), we get Equation (15):

$$\hat{P}_{H0}\in\left\{\left[\begin{bmatrix}0.944 & 0.023 & 0 & 0 & 0\\ 0.040 & 0.890 & 0.014 & 0 & 0\\ 0 & 0.017 & 0.926 & 0.012 & 0\\ 0 & 0 & 0.015 & 0.925 & 0.004\\ 0 & 0 & 0 & 0.001 & 0.981\end{bmatrix},\begin{bmatrix}0.977 & 0.056 & 0 & 0 & 0\\ 0.079 & 0.937 & 0.040 & 0 & 0\\ 0 & 0.045 & 0.962 & 0.037 & 0\\ 0 & 0 & 0.054 & 0.972 & 0.031\\ 0 & 0 & 0 & 0.019 & 0.999\end{bmatrix}\right]\right\}$$

Notice that the confidence region for the estimated transition matrix is not particularly large. We will come back to this point later. Now that we have calculated the confidence intervals for the estimated transition probabilities, we would like to project these intervals onto the space inhabited by the ergodic distribution. This is the issue to which we now turn.

3.2 Ergodic distribution

The estimated ergodic distribution is calculated as the left eigenvector corresponding to the unit eigenvalue of the estimated transition probability matrix (c.f. Equation 3). One might be tempted to apply this same line of reasoning in the calculation of the bounds to the confidence region for the ergodic distribution. This would mean calculating the lower bounds to the confidence intervals for the elements of the ergodic distribution by using the matrix of lower bounds to the confidence intervals for the transition probabilities to calculate the left eigenvector corresponding to the unit eigenvalue: $n^{\inf}(\infty)=n^{\inf}(\infty)P^{\inf}$. This is what is done in simple interval arithmetic. To add two intervals together, we add up the two lower bounds and add up the two upper bounds: $[a,b]+[c,d]=[a+c,b+d]$. To multiply two intervals together, we multiply the two lower bounds and multiply the two upper bounds: $[a,b]*[c,d]=[a*c,b*d]$. But taking a closer look at the matrix of lower bounds to the confidence intervals for the transition probabilities, one can notice that this matrix could never be an estimated probability matrix and therefore cannot be used to calculate an ergodic distribution. Indeed the rows do not sum to one, but to less than one, precisely because the matrix collects the lower bounds to the confidence intervals for the transition probabilities. The equivalent story can be told in terms of the upper bounds.

So, how can we identify the transition matrices within the confidence region that will generate the lowest and highest values for the different elements of the ergodic distribution? Hartfiel (1998) applies interval arithmetic to Markov chains in what is called the theory of Markov set-chains. In this context he presents the Hi-Lo Method, an algorithm to numerically calculate the bounds to a compact set of transition matrices at every step in its Markovian evolution. Since the rows of a transition matrix converge to the ergodic distribution as the Markov chain evolves towards its long-run distribution (c.f. Equation 3), applying the Hi-Lo Method until convergence is achieved will provide us with bounds for the ergodic distribution.

The Hi-LoMethod proceeds component by component, systematically identifying the lowest and highest values possible for each of the transition probabilities at every step in its path to convergence. Let us go through this algorithm for one of the transition probabilities. Consider the first entry in the transition matrix P at time t, $p_{11}(t)$. One period goes by, t becomes $t+1$, and P beomes P^2. The first entry in the transition matrix at time $t+1$ is obtained via matrix multiplication of the first column of P and the first row of P. This can be thought of as calculating the average of the column elements, weighted by the row elements. We would like to identify the smallest value that $p_{11}(t+1)$ can possibly take. Our starting point is $p_{11}^{\inf}(t)$, the first entry in the matrix of lower bounds $P^{\inf}$ at time t that is calculated using Equation (14). The first entry in the matrix of lower bounds at time $t+1$ is obtained via matrix multiplication of a column and a row. How to construct this column and this row? We want the smallest numbers possible subject to the constraint that the row must sum to one. Since there are no constraints concerning the column, we take the column containing the smallest values possible, i.e. the first column of $P^{\inf}$. As for the row, it needs to sum to one, which means that we cannot use the first row of $P^{\inf}$. The smaller some of the elements of the row are chosen to be, the bigger the other elements of the row are going to have to be. So we construct a row using the smallest elements from $P^{\inf}$ and the biggest elements from $P^{\sup}$, leaving one element to be calculated so that the row sums to one. These elements are arranged so as to minimize the weighted sum of the column. This means that the biggest elements of the row are positioned to weight the smallest elements of the column, leaving the smallest elements of the row to weight the biggest elements of the column. The same approach can be taken to calculate $p_{11}^{\sup}(t+1)$, the biggest value that $p_{11}(t+1)$ can possibly take. This methodology can then be applied to each of the transition probabilities at each time period until the rows of the transition matrix converge to the ergodic distribution. This will provide us with the bounds to the confidence intervals for each of the elements of the ergodic distribution.

Applying the Hi-Lo Method to the matrices of upper and lower bounds presented in (14), we get:

$$\hat{n}(\infty) \in \{[0.005 \quad 0.004 \quad 0.009 \quad 0.010 \quad 0.003], [0.710 \quad 0.449 \quad 0.464 \quad 0.443 \quad 0.932]\}. \quad (16)$$

Notice that the confidence region for the estimated twin-peaked distribution is inhabited by distributions of all shapes. For example, within this confidence region we find the uniform distribution $[0.20 \quad 0.20 \quad 0.20 \quad 0.20 \quad 0.20]$, the unimodal distribution $[0.10 \quad 0.20 \quad 0.40 \quad 0.20 \quad 0.10]$, and the trimodal distribution $[0.20 \quad 0.15 \quad 0.30 \quad 0.10 \quad 0.25]$. Coming back to the question posed at the beginning of this section, we can now answer that the twin-peaked result is not robust.

It is important to point out that the nonrobustness of the twin peaks result does not originate in the particularly large size of the confidence region for the estimated transition probability matrix (c.f. Equation 15), but rather in the extreme sensitivity of the ergodic distribution to perturbations to the estimated transition matrix. This extreme sensitivity was first noted by Reichlin (1999) in her discussion of Quah (1999). To illustrate this point, consider the following three examples in which the estimated transition matrix is only marginally perturbed and yet the corresponding ergodic distribution is significantly perturbed.

Increasing $\hat{p}_{23}$ by 1%, causes the ergodic distribution to become trimodal: $[0.25 \quad 0.16 \quad 0.20 \quad 0.14 \quad 0.25]$. Increasing $\hat{p}_{45}$ by 2%, shifts the mass in the ergodic distribution from the four left classes to the right peak: $[0.24 \quad 0.15 \quad 0.14 \quad 0.10 \quad 0.37]$. Increasing $\hat{p}_{54}$ by 1%, causes the ergodic distribution to become unimodal in state 1: $[0.34 \quad 0.21 \quad 0.19 \quad 0.14 \quad 0.12]$.

In the next section, the economic notion of elasticity is used to develop a measure of the fragility of the ergodic distribution in view of defining the empirical limitations of the simple Markov chain model.

4. Elasticity

In this section, the fragility of the ergodic distribution to perturbations to the estimated transition probability matrix is examined in analytical terms. The idea is to express the elements of the ergodic distribution as a function of the transition probabilities in order to calculate the elasticities of the former with respect to the latter. These elasticities then serve to describe certain characteristics of the ergodic function that is being fitted to the data and to explain the results presented in the previous section.

4.1 Ergodic function

In the application of Markov chain theory to the evolution of income distribution over time, the estimated transition probability matrix presents the following tri-diagonal structure:

$$P = \begin{matrix} a & b & 0 & 0 & 0 \\ c & d & f & 0 & 0 \\ 0 & g & h & j & 0 \\ 0 & 0 & k & l & m \\ 0 & 0 & 0 & n & o \end{matrix}, \text{ where } \begin{matrix} a+b=1 \\ c+d+f=1 \\ g+h+j=1 \\ k+l+m=1 \\ n+o=1 \end{matrix} \tag{17}$$

and $a,b,c,d,f,g,h,j,k,l,m,n,o > 0$. As usual, the ergodic distribution presents the following structure: $n(\infty) = v \quad w \quad x \quad y \quad z$ where $v+w+x+y+z=1$ and $v,w,x,y,z \geq 0$.

To express v, w, x, y, z as a function of $a, b, c, d, f, g, h, j, k, l, m, n, o$, the system of equations defining the ergodic distribution (i.e. $n(\infty) = n(\infty) \cdot P$) needs to be solved.

The following intermediate result is obtained:

$$v \quad w \quad x \quad y \quad z = \frac{c}{b}w \quad \frac{g}{f}x \quad \frac{k}{j}y \quad \frac{n}{m}z \quad z. \tag{18}$$

This result is interesting because it shows that the magnitude of any one element of the ergodic distribution relative to that of its right hand side neighbor is determined by the ratio of the probability of entry into, and the probability of exit from, the neighboring state. For example, if the probability of entering state 1 from state 2 (i.e. c) is double the probability of leaving state 1 to state 2 (i.e. b), then in the long run, the mass in state 1 (i.e. v) will be double the mass in state 2 (i.e. w). In other words, just by calculating the four ratios of off-diagonal elements, it is possible to establish the shape of the ergodic distribution.

Completely solving the system, the following result is obtained:

$$v \quad w \quad x \quad y \quad z = \frac{V}{\Sigma} \quad \frac{W}{\Sigma} \quad \frac{X}{\Sigma} \quad \frac{Y}{\Sigma} \quad \frac{Z}{\Sigma} \tag{19}$$

where $V = nkgc$, $W = nkgb$, $X = nkfb$, $Y = njfb$, $Z = mjfb$ and $\Sigma = V + W + X + Y + Z$. This result has a surprisingly simple interpretation. The portion of the population in any one state in the long run is just the product of the probabilities of transiting *towards* that state divided by the sum of the five possible products. For example, the portion of the population in state 1 in the long run (i.e. v) is the product of the probabilities of transiting from 5 to 4 (i.e. n), from 4 to 3 (i.e. k), from 3 to 2 (i.e. g), and from 2 to 1 (i.e. c) divided by the sum of the products of the probabilities of transiting towards 1, 2, 3, 4 and 5, respectively. Note the absence of the diagonal elements of the transition probability matrix from the function defining the elements of the ergodic distribution. This means that the ergodic distribution depends only upon observations of mobility and not upon observations of immobility.

Before using the ergodic function to calculate the elasticities of the elements of the ergodic distribution with respect to the transition probabilities, it is useful to explicitly state the conditions under which the transition matrix generates a twin-peaked ergodic distribution. For two modes located in classes 1 and 5, the following two conditions are necessary: $v > w \leftrightarrow c > b$ and $z > y \leftrightarrow m > n$. In other words, the probability of entering the state in which the mode is located must be superior to the the probability of leaving this state. To ensure that these two modes are the only two modes (i.e. that there is no mode located in state 3), either one or both of the following two conditions is necessary: $x \leq w \leftrightarrow f \leq g$ and/or $x \leq y \leftrightarrow k \leq j$. In other words, the probabilities of entering state 3 cannot simultaneously be superior to the probabilities of leaving state 3. Combining this interpretation of the off-diagonal elements with the interpretation of the ratios of the off-diagonal elements, it is possible just by looking at the transition matrix to determine the existence and relative magnitude of twin peaks in the ergodic distribution.

4.2 Elasticity

Now let us calculate the elasiticities of the elements of the ergodic distribution with respect to the transition probabilities. Since all the elements concerned are probabilities, this elasticity is just the slope of the ergodic function of interest. Since each element of the ergodic distribution is a function of the eight off-diagonal elements of the transition matrix, and since the ergodic distribution is made up of five elements, there are 40 elasticities to calculate for each ergodic distribution. The general expressions for these 40 elasticities are presented in the Figure 8, and the values taken by these expressions in each of the five cases considered are presented in Figure 9. These elasticities reveal two interesting characteristics of the ergodic function.

First, the elasticities of the elements of the ergodic distribution with respect to the transition probabilities tend to increase as the off-diagonal elements of the transition matrix decrease. For example, a marginal perturbation (+/- 1%) to c when $b = c = f = g = j = k = m = n = 0.03$ can either create a peak or a trough in the poorest class of the long run income distribution (i.e. $n(\infty) = [0.25 \quad 0.19 \quad 0.19 \quad 0.19 \quad 0.19]$ or $n(\infty) = [0.14 \quad 0.21 \quad 0.21 \quad 0.21 \quad 0.21]$), whereas the same marginal perturbation to c when $b = c = f = g = j = k = m = n = 0.33$ barely affects the uniformity of the poorer classes of the long run income distribution (i.e. $n(\infty) = [0.205 \quad 0.199 \quad 0.199 \quad 0.199 \quad 0.199]$ or

$n(\infty) = [0.195 \quad 0.201 \quad 0.201 \quad 0.201 \quad 0.201]$). Equation (18) is useful in understanding why this happens to be the case. When the off-diagonal elements of the transition matrix are small, a small absolute difference (e.g. $c - b = 0.05 - 0.01 = 0.04$) can be compatible with a large relative difference (e.g. $c / b = 0.05 / 0.01 = 5$), whereas when these elements are larger, the same absolute difference (e.g. $c - b = 0.35 - 0.31 = 0.04$) corresponds to a smaller relative difference (e.g. $c / b = 0.35 / 0.31 = 1.13$). Since it is the relative differences that determine the shape of the ergodic distribution, the smaller the off-diagonal elements of the transition matrix are, the more sensitive the shape is to marginal perturbations to these off-diagonal elements.

Second, the elasticities of *v* and *z* with respect to the transition probabilities are higher than the elasticities of *w*, *x* and *y* with respect to the transition probabilities. Intuitively, the relatively high elasticity of the endpoints of the ergodic distribution can be explained by the fact that whereas the three middle states all have two exits (i.e. one to a lower state and one to a higher state), the two endpoints only have one exit. So, whereas blocking one of these single exits necessarily leads to a pile up in one of the endpoints in the long run, blocking one of the other exits leads to a pile up that could potentially be distributed amongst all the states behind the blocked exit via the unblocked exit. Another way of thinking about this idea is by counting the number of constraints on the elements of the transition probability matrix required to generate a mode in the different states of the ergodic distribution. The generation of a mode in a particular state requires that the entry probabilities be superior to the exit probabilities. Since the states 1 and 5 only have one point of entry and of exit, only one condition needs to be fulfilled to ensure a mode in one of these states in the long run (i.e. $c > b$ for a mode in state 1 and $m > n$ for a mode in state 5). States 2, 3 and 4, on the other hand, each have two points of entry and of exit and therefore two conditions need to be fulfilled to ensure a mode in one of these states in the long run (i.e. $b > c$ and $g > f$ for a mode in state 2, $f > g$ and $k > j$ for a mode in state 3, and $j > k$ and $n > m$ for a mode in state 4). In sum, with such a functional form, it is easier to generate a mode in states 1 and 5 than in any of the other states, especially if the off-diagonal elements of the transition matrix are particularly small.

4.3 Application

Finally, we examine the insights provided by these two observations concerning the ergodic function into the results presented in Section 3. Looking at the ergodic distribution calculated from the unfiltered data (see Equation 6), we observe that sizeable twin peaks are present, and we recall that these twin peaks are not statistically significant. Why? As for the existence of the twin peaks, the conditions for modes located in states 1 and 5 of the ergodic distribution are fulfilled (i.e. $c = 0.0594 > b = 0.0396$ and $m = 0.0173 > n = 0.0103$), and the conditions for modes located in states 2, 3 or 4 are not fulfilled. As for the magnitudes of the twin peaks, the $c/b = 1.5$ and $m/n = 1.7$ ratios are relatively high. Note that $c/b = 1.5$ even though $c - b = 0.0198$, and that $m/n = 1.7$ even though $m - n = 0.0070$. In other words, the sizeable twin peaks result from minute differences between the related off-diagonal transition probabilities.

As for the lack of statistical significance of the twin peaks, the off-diagonal elements of the transition matrix are relatively small, the elasticities of the elements of the ergodic distribution are relatively high, and the ergodic distribution is relatively sensitive to variations in the elements of the transition matrix. In particular, the mode in state 1 of the ergodic distribution exists because $c = 0.0594 > b = 0.0396$. *Relative to the uniform distribution*, this difference between c and b represents either 11 observations too few out of the 555 observations made in state 1, or 11 observations too many out of the 556 observations made in state 2. The case surrounding the mode in state 5 is even more impressive. The mode in state 5 of the ergodic distribution exists because $m = 0.0173 > n = 0.0103$. *Relative to the uniform distribution*, this difference between m and n represents either 1 observation too few out of the 347 observations made in state 4, or 1 observation too many out of the 487 observations made in state 5. For the mode in state 5 to be significant, no error could be tolerated and the confidence region would have to be a point.

Comparing the results from the estimation with no filter to the results from the estimations with increasingly fine filters, the observations made above become all the more true. Indeed, the filters render the transition matrices less and less mobile by definition, thus progressively reducing the already small off-diagonal elements of the transition matrix and progressively increasing the already large elasticities of the elements of the ergodic distribution relative to the off-diagonal elements of the transition matrices. This extreme fragility of the elements of these ergodic distributions to variations in the transition probabilities is illustrated by the impact of the third filter on the ergodic distribution (see Equation 6). This filter discards three of the observed transitions into state 1 and two of the observed transitions out of state 1 and in consequence the mode in state 1 of the ergodic distribution rises from around 40% to around 50%.

Observing the evolution of the results as increasingly fine filters are applied, we see a rapid degeneration of the irreducible Markov chain model into an absorbing Markov chain model. The results of the fourth filter indicate that of the 487 observations in state 5, only 2 constitute transitions, and that of the 555 observations in state 1, only 6 constitute transitions (see Figure 7). Thus it would not be very farfetched to fit the data with a model treating states 1 and 5 as absorbing states. In a model with two absorbing states, all observations end up in one of the two states (i.e. twin peaks), and the distribution between these two states depends not only upon the transition probabilities in the transient class, but also upon the initial distribution of the observations. Calculating this long run distribution between the two states, we obtain that if the underlying processes were to continue in the long run, 66% of the countries would end up very poor and 34% very rich.

In sum, not only does the application of the simple Markov chain model to data on income distribution impose a very particular functional form to the long run distribution of the data, but the data is also such that it situates the analysis in the most fragile region of this very particular functional form, in a region where the irreducible Markov chain model is rapidly degenerating into an absorbing Markov chain model.

5. Concluding remarks

This paper makes two contributions to the existing literature on empirical applications of the simple Markov chain model. First, confidence intervals for the transition probabilities are analytically derived and confidence intervals for the ergodic probabilities are numerically calculated. These are the basic elements of statistical inference that should constitute an integral part of any empirical analysis. Second, the notion of the elasticity of the ergodic probabilities with respect to the transition probabilities is developed and quantified. This is a tool that researchers can use to assess whether or not the simple Markov chain model should be used to model their data.

Two lessons emerge from this paper. First, when Markov chain models are fitted to continuous data, a bias towards excess mobility is introduced into the estimated transition matrix. Second, when the estimated transition matrix of a simple Markov chain model presents a certain type of high immobility, the corresponding ergodic distribution is characterized by an extreme fragility of a very particular sort.

6. References

Anderson, T.W., and L.A. Goodman, 1957, Statistical inference about Markov chains, Annals of Mathematical Statistics 28, 89-110.

Hartfiel, D.J., 1998, Markov Set-Chains, Springer, Berlin.

Quah, D., 1993a, Empirical cross-section dynamics in economic growth, European Economic Review 37, 426-434.

Quah, D., 1993b, Galton’s fallacy and tests of the convergence hypothesis, Scandinavian Journal of Economics 95, 427-443.

Quah, D., 1996, Empirics for economic growth and convergence, European Economic Review 40, 1353-1375.

Quah, D., 1999, Convergence as distribution dynamics In Baldwin, R., Cohen, D., Sapir, A., and A. Venables (Eds.), Regionalism and the Global Economy, Cambridge University Press, Cambri

Reichlin, L., 1999, Convergence as distribution dynamics: discussion In Baldwin, R., Cohen, D., Sapir, A., and A. Venables (Eds.), Regionalism and the Global Economy, Cambridge University Press, Cambridge.

Summers, R. and A. Heston, 1991, The Penn World Table (Mark 5): an expanded set of international comparisons, 1950-1988, The Quarterly Journal of Economics 106, 327-368.

Figure 3: Business cycle type fluctuations about the frontier between states 1 and 2

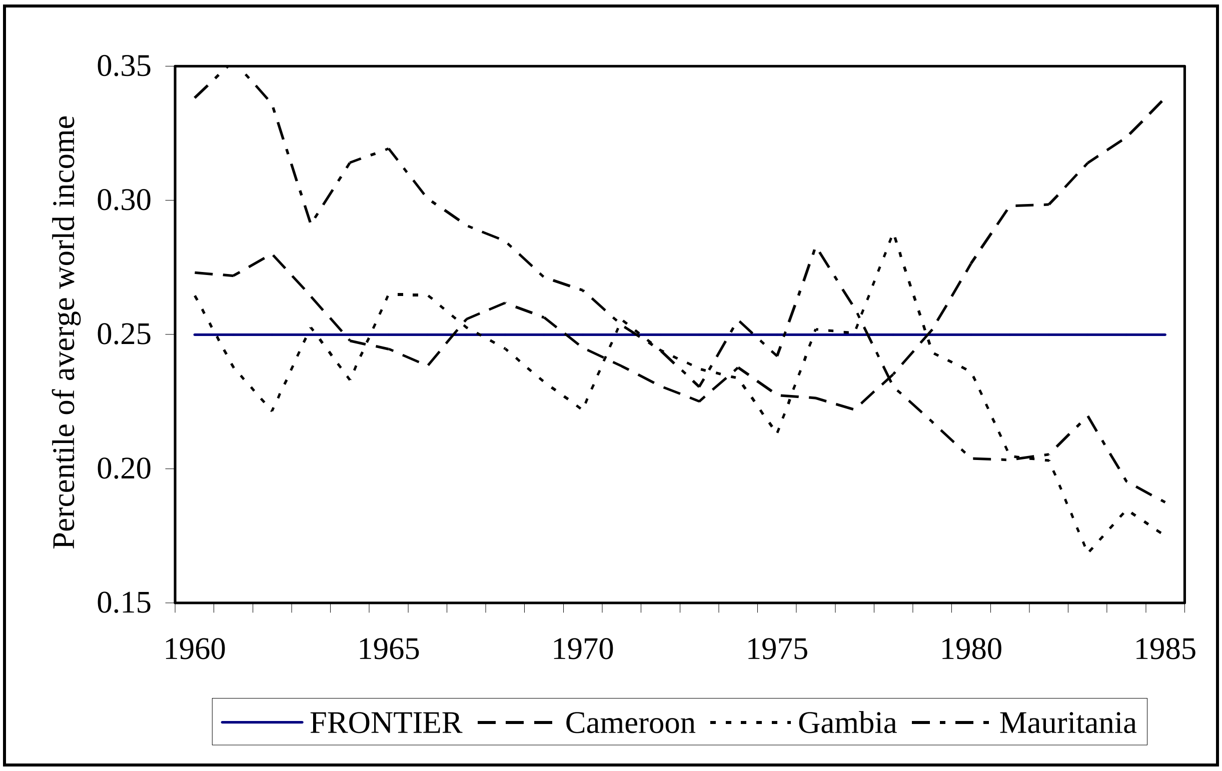


Figure 4: Business cycle type fluctuations about the frontier between states 2 and 3

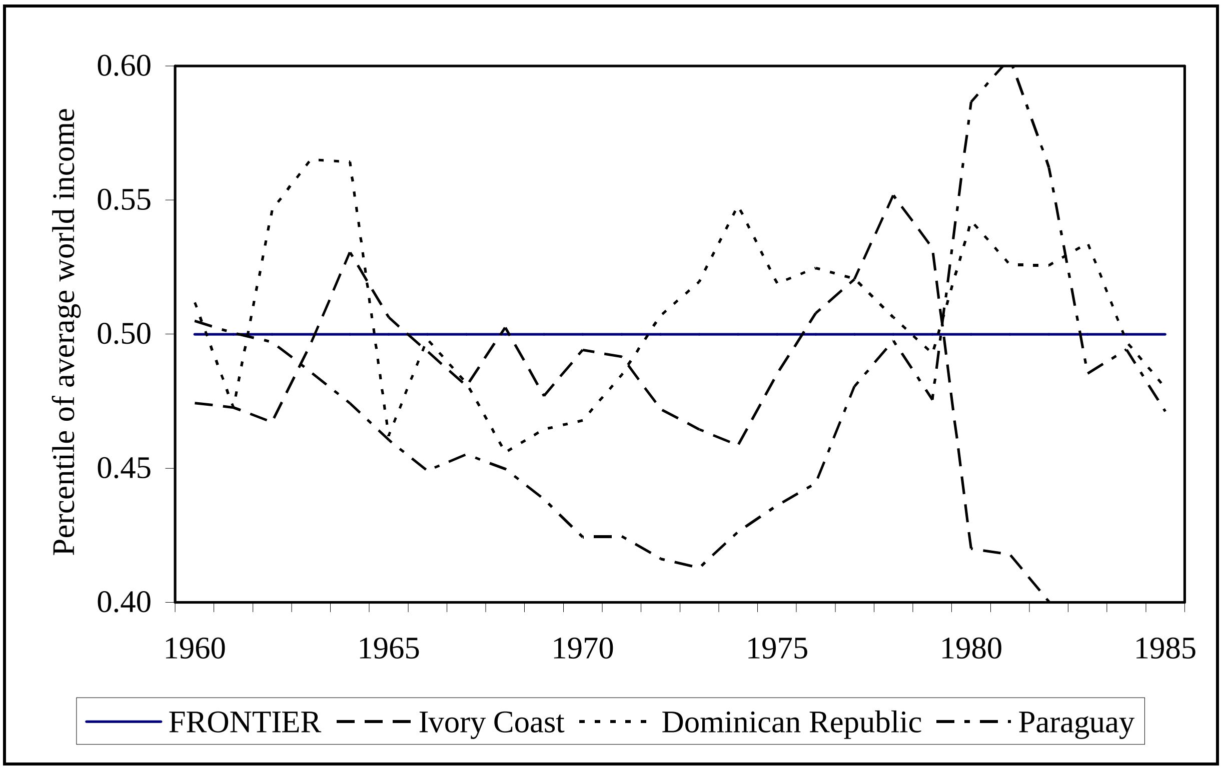


Figure 5: Business cycle type fluctuations about the frontier between states 3 and 4

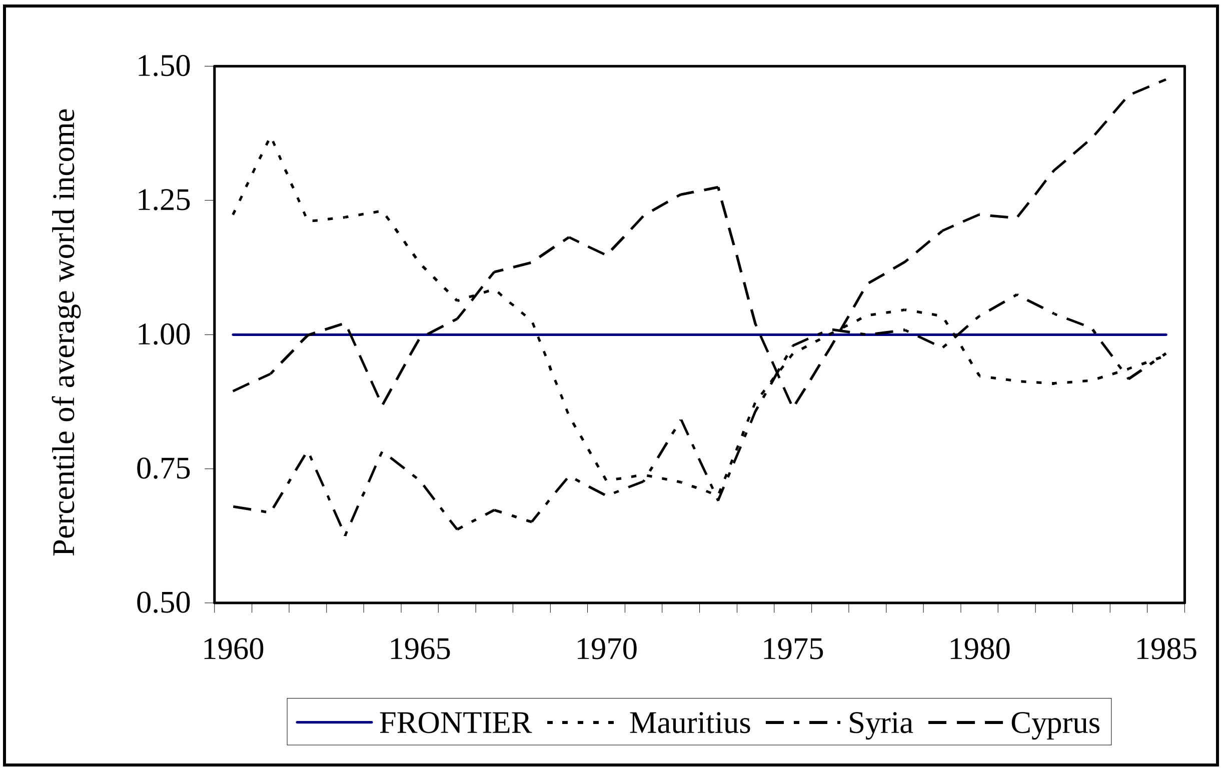


Figure 6: Business cycle type fluctuations about the frontier between states 4 and 5

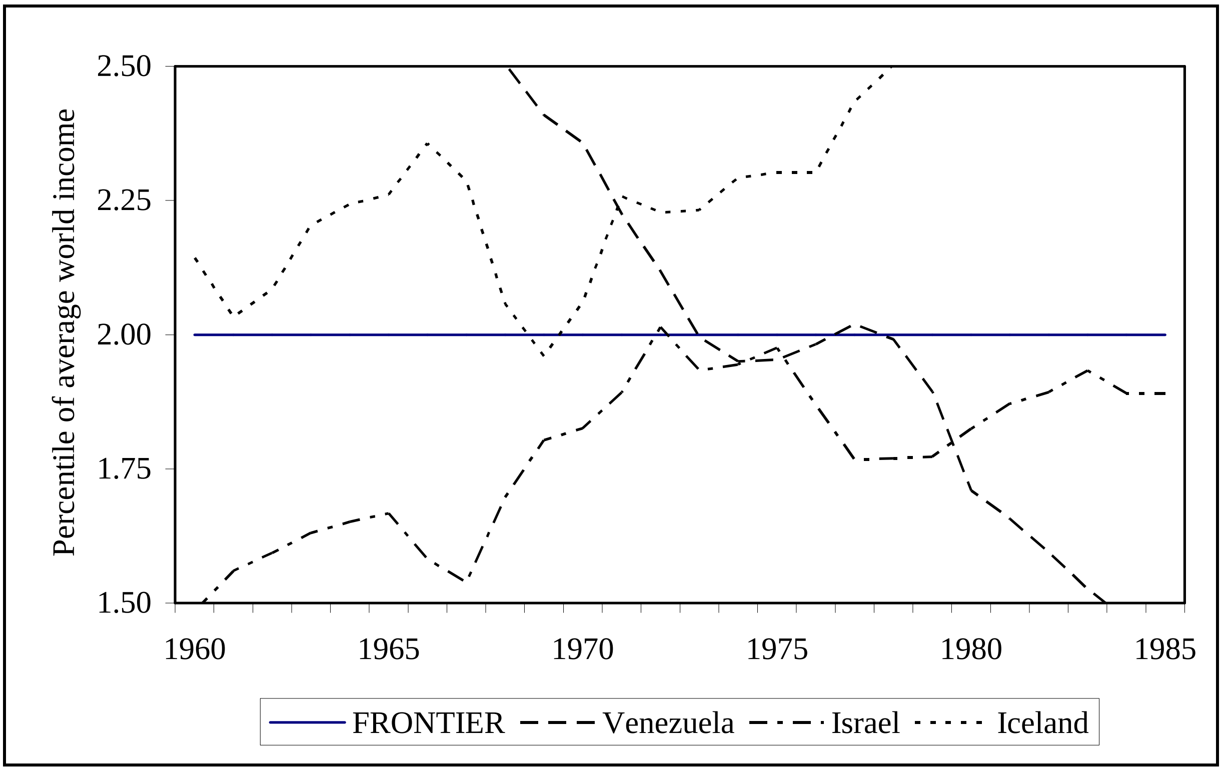


Figure 7: Observed transition matrices

$$N_0 = \begin{matrix} 533 & 22 & 0 & 0 & 0 \\ 33 & 508 & 15 & 0 & 0 \\ 0 & 19 & 574 & 15 & 0 \\ 0 & 0 & 12 & 329 & 6 \\ 0 & 0 & 0 & 5 & 482 \end{matrix}$$

$$N_1 = \begin{matrix} 544 & 11 & 0 & 0 & 0 \\ 28 & 516 & 12 & 0 & 0 \\ 0 & 16 & 581 & 11 & 0 \\ 0 & 0 & 12 & 330 & 5 \\ 0 & 0 & 0 & 3 & 484 \end{matrix}$$

$$N_2 = \begin{matrix} 547 & 8 & 0 & 0 & 0 \\ 25 & 520 & 11 & 0 & 0 \\ 0 & 13 & 584 & 11 & 0 \\ 0 & 0 & 9 & 334 & 4 \\ 0 & 0 & 0 & 2 & 485 \end{matrix}$$

$$N_3 = \begin{matrix} 549 & 6 & 0 & 0 & 0 \\ 22 & 526 & 8 & 0 & 0 \\ 0 & 12 & 587 & 9 & 0 \\ 0 & 0 & 8 & 335 & 4 \\ 0 & 0 & 0 & 2 & 485 \end{matrix}$$

$$N_4 = \begin{matrix} 549 & 6 & 0 & 0 & 0 \\ 20 & 529 & 7 & 0 & 0 \\ 0 & 12 & 587 & 9 & 0 \\ 0 & 0 & 7 & 336 & 4 \\ 0 & 0 & 0 & 2 & 485 \end{matrix}$$

Figure 8: General expressions for the elasticities of the elements of the ergodic distribution with respect to the transition probabilities

$$e_{vb}=\frac{-(nkgc)(nkg+fnk+jfn+mjf)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{vc}=\frac{(nkg)(bnkg+fbnk+jfbn+mjfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{wb}=\frac{(nkg)(nkgc)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{wc}=\frac{-(bnkg)(nkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{xb}=\frac{(fnk)(nkgc)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{xc}=\frac{-(fbnk)(nkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{yb}=\frac{(jfn)(nkgc)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{yc}=\frac{-(jfbn)(nkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{zb}=\frac{-(mjf)(nkgc)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{zc}=\frac{-(mjfb)(nkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{vf}=\frac{-(nkgc)(bnk+jbn+mjb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{vg}=\frac{(nkc)(fbnk+jfbn+mjfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{wf}=\frac{-(bnkg)(bnk+jbn+mjb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{wg}=\frac{(bnk)(fbnk+jfbn+mjfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{xf}=\frac{(bnk)(nkgc+bnkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{xg}=\frac{-(fbnk)(nkc+bnk)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{yf}=\frac{(jbn)(nkgc+bnkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{yg}=\frac{-(jfbn)(nkc+bnk)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{zf}=\frac{(mjb)(nkgc+bnkg)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{zg}=\frac{-(mjfb)(nkc+bnk)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{vj}=\frac{-(nkgc)(fbn+mfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{vk}=\frac{(ngc)(jfbn+mjfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{wj}=\frac{-(bnkg)(fbn+mfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{wk}=\frac{(bng)(jfbn+mjfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{xj}=\frac{-(fbnk)(fbn+mfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{xk}=\frac{(fbn)(jfbn+mjfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{yj}=\frac{(fbn)(nkgc+bnkg+fbnk)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{yk}=\frac{-(jfbn)(ngc+bng+fbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{zj}=\frac{(mfb)(nkgc+bnkg+fbnk)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{zk}=\frac{-(mjfb)(ngc+bng+fbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{vj}=\frac{-(nkgc)(fbn+jfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{vk}=\frac{(kgc)(jfbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{wj}=\frac{-(bnkg)(jfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{wk}=\frac{(bkg)(jfbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{xj}=\frac{-(fbnk)(jfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{xk}=\frac{(fbk)(jfbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{yj}=\frac{-(jfbn)(jfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{yk}=\frac{(jfb)(jfbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

$$e_{zj}=\frac{(jfb)(nkgc+bnkg+fbnk+jfbn)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2} \qquad e_{zk}=\frac{-(mjfb)(kgc+bkg+fbk+jfb)}{(nkgc+bnkg+fbnk+jfbn+mjfb)^2}$$

Figure 9: Elasticities of the elements of the ergodic distribution with respect to the transition probabilities

$$e = \begin{matrix} e_{vb} & e_{vc} & e_{vf} & e_{vg} & e_{vj} & e_{vk} & e_{vm} & e_{vn} \\ e_{wb} & e_{wc} & e_{wf} & e_{wg} & e_{wj} & e_{wk} & e_{wm} & e_{wn} \\ e_{xb} & e_{xc} & e_{xf} & e_{xg} & e_{xj} & e_{xk} & e_{xm} & e_{xn} \\ e_{yb} & e_{yc} & e_{yf} & e_{yg} & e_{yj} & e_{yk} & e_{ym} & e_{yn} \\ e_{zb} & e_{zc} & e_{zf} & e_{zg} & e_{zj} & e_{zk} & e_{zm} & e_{zn} \end{matrix}$$

$$e_{H0} = \begin{matrix} -5.5 & 3.5 & -5.5 & 5.0 & -4.2 & 3.1 & -3.7 & 6.5 \\ 1.5 & -0.9 & -3.5 & 3.1 & -2.7 & 2.0 & -2.4 & 4.1 \\ 1.3 & -0.8 & 3.0 & -2.7 & -2.4 & 1.8 & -2.1 & 3.7 \\ 1.0 & -0.6 & 2.2 & -2.0 & 3.4 & -2.5 & -1.6 & 2.7 \\ 1.7 & -1.1 & 3.8 & -3.4 & 5.9 & -2.5 & 9.8 & -17.0 \end{matrix}$$

$$e_{H1} = \begin{matrix} -12.3 & 4.8 & -7.5 & 6.6 & -6.3 & 3.4 & -5.6 & 13.5 \\ 3.5 & -1.4 & -3.0 & 2.6 & -2.5 & 1.3 & -2.2 & 5.3 \\ 3.1 & -1.2 & 3.7 & -3.2 & -2.2 & 1.2 & -1.9 & 4.6 \\ 1.7 & -0.7 & 2.0 & -1.8 & 3.2 & -1.7 & -1.1 & 2.5 \\ 4.0 & -1.6 & 4.8 & -4.2 & 7.7 & -1.7 & 10.8 & -25.9 \end{matrix}$$

$$e_{H2} = \begin{matrix} -17.6 & 5.4 & -9.1 & 9.0 & -7.8 & 5.7 & -9.2 & 26.5 \\ 3.7 & -1.1 & -2.8 & 2.7 & -2.4 & 1.7 & -2.8 & 8.0 \\ 3.7 & -1.1 & 3.1 & -3.1 & -2.4 & 1.7 & -2.8 & 8.0 \\ 2.7 & -0.8 & 2.3 & -2.2 & 3.3 & -2.4 & -2.0 & 5.8 \\ 7.6 & -2.3 & 6.5 & -6.4 & 9.3 & -2.4 & 16.8 & -48.2 \end{matrix}$$

$$e_{H3} = \begin{matrix} -25.2 & 6.2 & -12.2 & 9.4 & -9.0 & 6.1 & -8.7 & 24.9 \\ 6.7 & -1.6 & -3.0 & 2.3 & -2.2 & 1.5 & -2.1 & 6.1 \\ 5.1 & -1.2 & 4.2 & -3.3 & -1.7 & 1.2 & -1.6 & 4.7 \\ 3.5 & -0.9 & 2.8 & -2.2 & 3.4 & -2.3 & -1.1 & 3.2 \\ 9.9 & -2.4 & 8.2 & -6.3 & 9.7 & -2.3 & 13.5 & -38.8 \end{matrix}$$

$$e_{H4} = \begin{matrix} -25.2 & 6.7 & -14.1 & 9.3 & -9.3 & 7.1 & -8.8 & 25.4 \\ 6.7 & -1.9 & -3.7 & 2.5 & -2.5 & 1.9 & -2.4 & 6.8 \\ 4.6 & -1.2 & 4.5 & -3.0 & -1.6 & 1.2 & -1.6 & 4.5 \\ 3.5 & -0.9 & 3.5 & -2.3 & 3.5 & -2.6 & -1.2 & 3.4 \\ 10.1 & -2.7 & 9.9 & -6.5 & 9.9 & -2.6 & 13.9 & -40.0 \end{matrix}$$

Chapter 2

Fuzzifying the Cross-Country Income Convergence Debate

Abstract: The simplicity with which short-run dynamics are translated into long-run tendencies in the simple Markov chain model has lured social scientists into applying this discrete model to continuous data. In the context of the cross-country convergence debate, this application leads to biased estimates of the short-run dynamics (i.e. transition probability matrix) and non-robust estimates of the long-run tendencies (i.e. ergodic distribution); the model chosen specifically because of the insights provided into mobility, no longer supplies robust results on mobility. This chapter examines the causes underlying this breakdown of the simple Markov chain model and proposes a simple solution that uses notions imported from the domain of fuzzy logic to adapt the simple Markov chain model to continuous data.

Keywords: Convergence; Filters; Fuzzy logic; Income distribution; Markov chains; Twin peaks.

JEL classification: C49; F02.

1. Introduction

The simple Markov chain model[3] has been widely used in the social sciences to study the phenomenon of mobility. In the early days of this literature, empirical applications were mainly limited to different aspects of social mobility (Rogoff 1953; Glass and Hall 1954; Blumen, Kogan and McCarthy 1955; Prais 1955; McFarland 1970; Singer and Spilerman 1974; Shorrocks 1976; Lee, Judge and Zellner 1977). Exceptions include Anderson (1954) who studies changes in voter attitude during the 1940 American presidential elections, Adelman (1958) who studies the size distribution of firms and Telser (1962) who studies the choice of cigarette brands. Quah (1993a, 1993b, 1996a) revives economists' interest in the simple Markov chain model with his application of distribution dynamics to the cross-country income convergence issue. Examples from the branch in the literature inspired by Quah's research include Villaverde and Sanchez-Robles (2001) in their study of the Spanish regional dynamics of per capita income, Fiaschi and Lavezzi (2004) in their study of the shape of the growth process for a cross-section of countries in the European Union and in the world, and Proudman, Redding and Bianchi (1998) in their study of the relationship between international openness in trade and economic growth. Recently, applications of the simple Markov chain model have been carried out in domains as diverse as agricultural economics (Temel and Alberson (2000) study the evolutions of the size and productivity distributions of farms in the US; Temel and Alberson (2001) study convergence in hired farm wages in United States counties), environmental economics (Skaggs and Ghosh (1999) study the changes in wind-based soil erosion rates over time), labor economics (Buchinsky and Hunt (1996) study wage mobility in the United States; Constant and Zimmerman (2003) study the dynamics of repeat migration in Germany), and even the economics of sports (Koop (2001) studies mobility across the performance distribution in American Major League Baseball).

.

The prime attraction in this Markovian approach lies in the simplicity with which short-run dynamics are characterized and long-run tendencies calculated.[4] In the previous chapter, I discussed the potential fragility of the estimated ergodic distribution (c.f. long-run tendencies), and I used the cross-country income convergence application to illustrate the

[3] The term *simple Markov chain model* refers to the representation of data via a first-order, discrete time, discrete state-space Markovian process.

[4] In the simple Markov chain model, the transition probability matrix summarizes the information on short-run dynamics and the ergodic distribution summarizes the information on long-run tendencies.

discussion. In this chapter, I will discuss the potential biasedness of the estimated transition matrix (c.f. short-run dynamics), and I will continue to use the same application to illustrate the discussion.

Quah uses the simple Markov chain model to paint a new picture of the world in which the rich and the poor are diverging to form 'twin peaks'. In Chapter 1, I contest the robustness of this conclusion, questioning the suitability of this model to analyse the cross-country convergence issue. More specifically, I show that the ergodic distribution is so exceedingly sensitive to marginal perturbations to the estimated transition probability matrix that all sorts of different shaped ergodic distributions are compatible with the transition probability matrices inhabiting the confidence region of the estimated transition probability matrix. In this chapter, I complete my critique of the application of the simple Markov chain model to the cross-country convergence issue. More specifically, I demonstrate the inadequacy of the estimated transition probability matrix in providing accurate information on short-run dynamics. So, in the context of the cross-country convergence issue, the simple Markov chain model provides a biased characterisation of the short-run dynamics and a fragile representation of the long-run tendencies. Other related problems with this Markovian approach have also been raised in the literature (Aghevli and Mehran 1981; Davies and Shorrocks 1989; Mattsson and Thorburn 1989; Fingleton 1997; Magrini 1999; Reichlin 1999; Landon-Lane and Quinn 2000; Bulli 2001; Kremer, Onatski and Stock 2001; Pearlman 2003), but efforts to resolve these problems have been piecemeal, and as such, unsatisfactory (for an exception, see Rummel 2005). In my view, this riddle has not yet been resolved because the problem has not been properly defined.

As most econometric modelling problems, this problem has two separate parts: model specification and model identification. If the specification is wrong, the identification is meaningless. In the literature on the cross-country income convergence issue, model specification is simply assumed away, and research is focused on the identification issue. But it is a well known and often acknowledged fact that it is wrong to just assume that the Markov property holds for a given time series of observations. This oversight is crucial because it leads to mistakes in the evaluation of the model identification. In the literature, the quality of the discrete Markov model is evaluated by comparing it with the continuous Markov model, but if the time discretization of the underlying continuous process that is arbitrarily imposed by the annual deadlines of national statistical offices is not Markovian, then the discrete and the continuous representations of the space dimension of the process are both wrong.

Even if I abstract away from this problem of model specification and focus on the model identification aspect of the cross-country convergence issue, I find that the approach taken in the literature is surprisingly unsatisfactory. Indeed, in this context space-discretization is just non-parametric estimation of a two-dimensional density, and yet the space-discretization carried out in the literature ignores even the simplest of the recommendations made in the vast literature on density estimation.

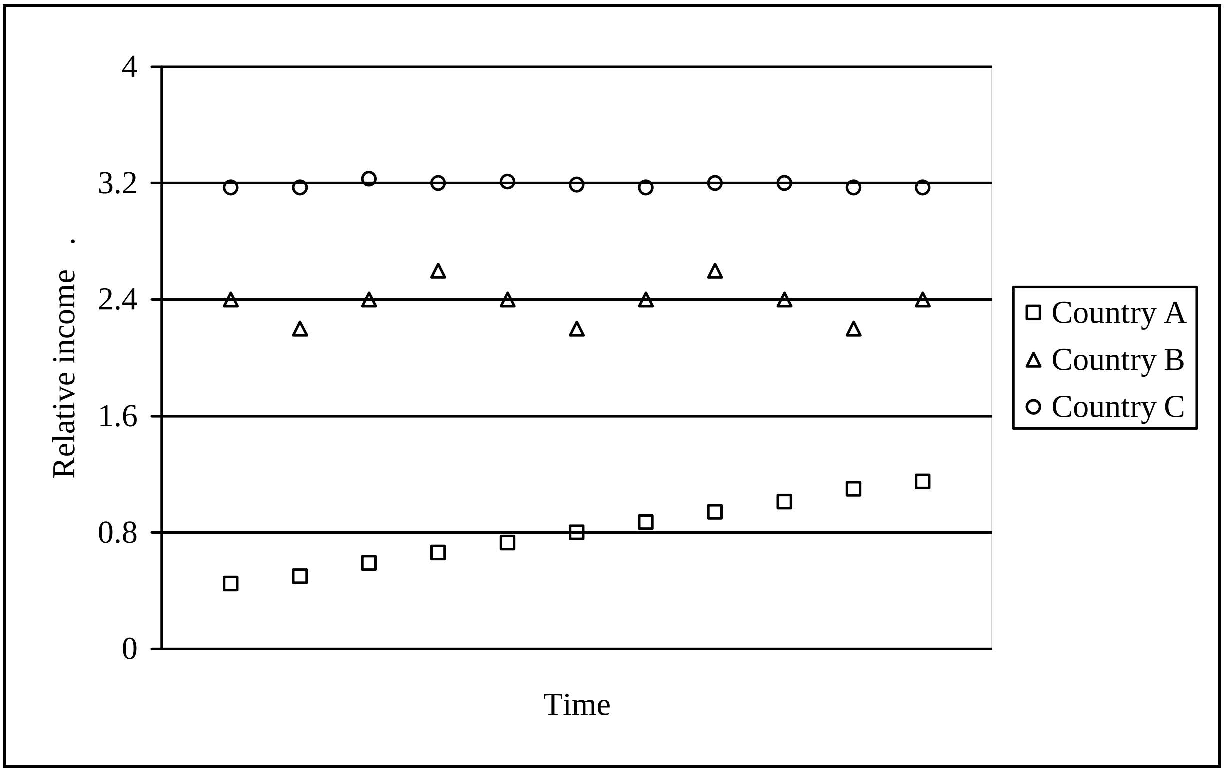
In this chapter, I propose an improvement upon past practice of space-discretization, one which resolves the problem of the fragility of the estimated ergodic distribution. This improvement presents the additional advantage of also resolving the yet to be defined problem of the biasedness of the estimation transition probability matrix.

In the application of the simple Markov chain model to panel data on income, the estimated transition probability matrix is used to extract information concerning the mobility of countries within the distribution of incomes from the dataset. This information is camouflaged by two sources of noise. In theory, the long-run distribution is independent of short-run noise; in practice, however, this is no longer the case because the sample is finite. Moreover, because of the exceedingly high sensitivity of the ergodic distribution to marginal perturbations to the estimated transition probability matrix, the short-run noise does not even have to be sizeable to significantly contaminate the long-run distribution.

The first source of noise results from using continuous data to estimate a categorical model (i.e. from translating continuous data into discrete data by defining income class frontiers). In the simple Markov chain model, transitions represent mobility; in reality, however, transitions can occur for two reasons. Transitions can result from higher or lower than average world growth in a country. In Figure 1, country A’s growth process propels it through the cross-country income distribution in an upward manner, while country B’s business cycle steers it through the cross-country income distribution in a cyclical fashion. This is what we call mobility and this is what we would like to measure. Transitions can also result from marginal fluctuations in a country's relative income when the level of relative income is situated very close to one defining a frontier between classes. In Figure 1, country

C follows a surprisingly stable path through the cross-country income distribution. This is clearly not mobility and since such transitions are included in the calculation of mobility, it is necessary to correct the estimated transition probability matrix for this bias.

Figure 1: Transitions resulting from growth (A), business cycles (B), and marginal fluctuations (C)



The second source of noise is generated by inaccuracies in the data. Two aspects of these inaccuracies are of particular interest. First, data inaccuracy exists. A given fluctuation in the context of a constantly-evolving series on estimates of income should not necessarily be treated in the same manner as the same fluctuation in the context of an upwardly-evolving series of estimates on income. A distinction needs to be made between short-run dynamics and short-run noise. Second, there are huge differentials in data inaccuracy. Heston, Summers and Aten (2006) and Summers and Heston (1984, 1991, 1994, 2002) include quality grades for the data provided in the Penn World Tables. A series of badly-graded estimates of income should not be treated in the same manner as a series of well-graded estimates of income. A distinction needs to be made between good-quality data and bad-quality data. The robustness of the results provided by the simple Markov chain model significantly increases when these three considerations are taken into account.

In sum, the simplicity with which short-run dynamics are translated into long-run tendencies in the simple Markov chain model has convinced researchers to force this categorical model onto continuous data, and paradoxically enough, this forcing (enhanced by the data inaccuracy issue) biases estimates of the short-run dynamics and fragilizes estimates of the long-run tendencies. In other words, the model chosen specifically because of the insights provided into mobility, no longer supplies robust results on mobility. The obvious solution to this problem is to use a continuous framework when working with continuous data, as with the non-parametric techniques applied by Quah (1996a, 1997), Bianchi (1997) and Johnson (2000). Another solution is to perform a rigorous discretization of the continuous data before applying the categorical model, as with the regenerative sampling carried out by Bulli (2001). These solutions, although valid, compromise the simplicity initially attracting researchers to the simple Markov chain model. This chapter provides such a simple solution, using notions imported from the domain of fuzzy logic to adapt the simple Markov chain model to continuous data.[5]

The remainder of the chapter is organized as follows. In Section 2, the cross-country income convergence issue is redefined in order to highlight its two main features, time-discretization and space-discretization. Discrete and continuous space-discretization are discussed in Sections 2.1 and 2.2, respectively. Time discretization is dealt with in Section 2.3. Section 3 illustrates the problem posed by the application of the simple Markov chain model to panel data on income, focusing on the bias present in the estimate of the short-run dynamics. In the Section 3.1, the consequences of the arbitrary discretization of the continuous data for the estimated transition probability matrix are analyzed. In Section 3.2, the implications of the two types of data inaccuracy for the estimate of the transition probability matrix are examined. Sections 4 and 5 are devoted to presenting a simple solution permitting the application of the simple Markov chain model to continuous data. In Section 4, fuzzification of the income class frontiers and of the income observations is carried out. In Section 5, a selective filter is applied to the data, eliminating the short-run noise while

[5] This solution has nothing to do with *fuzzy Markov chains* (Avrachenkov and Sanchez 2002).

retaining the short-run dynamics. Section 6 presents the empirical results and Section 7 concludes.

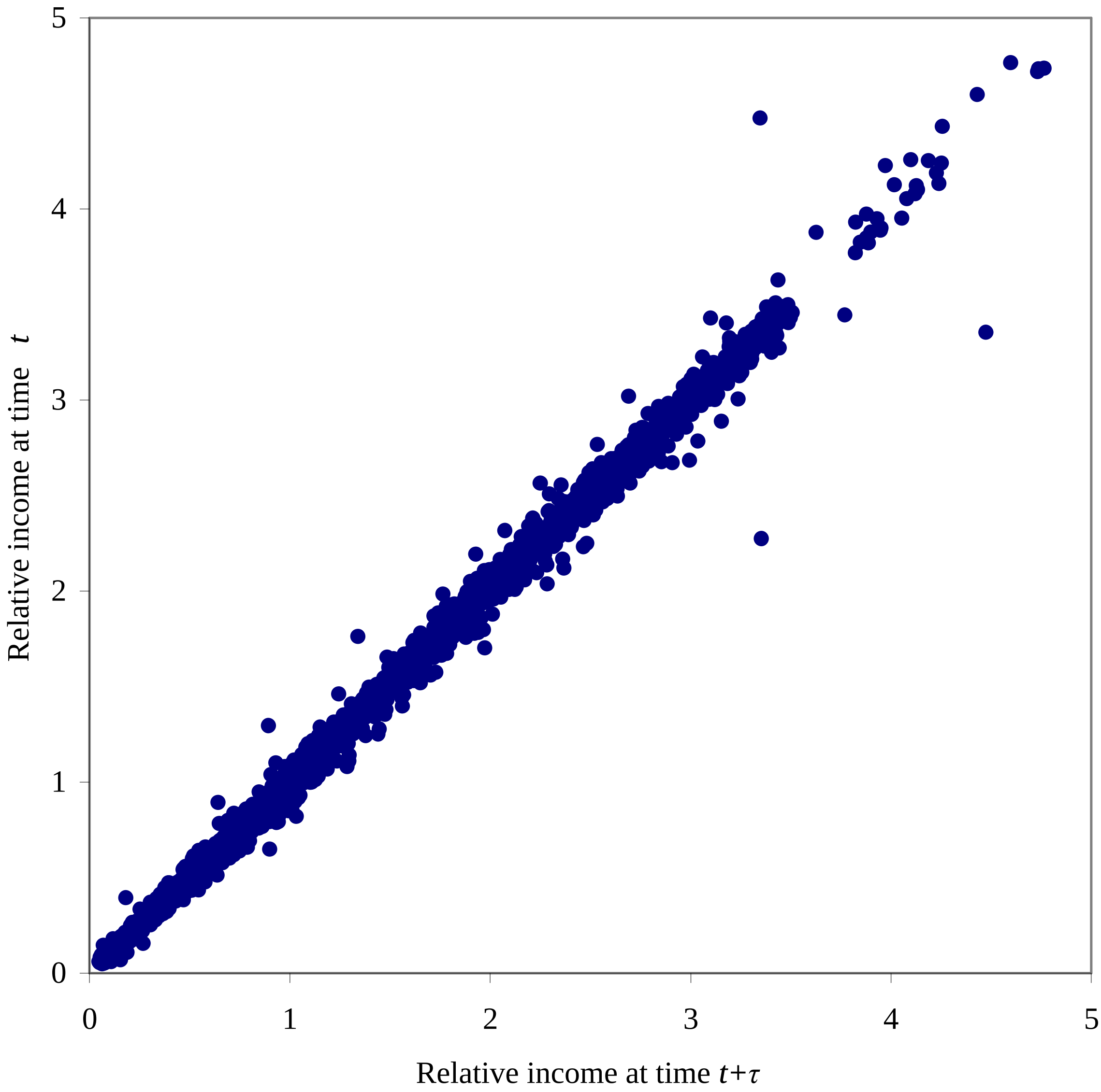
2. Redefinition of the cross-country income convergence issue

Countries follow different paths over time through the distribution of income across countries. These continuous space and time processes of income mobility are observed at discrete points in time $t = 0, t, 2t, ..., T$ when they take values belonging to the continuous space of relative incomes *S*. The data used is the Laspeyres index of annual real per capita income from the Heston, Summers and Aten (2006) Penn World Tables version 6.2 for the 98 countries with data available for the 1960-03 time period.[6] The variable used to evaluate mobility is relative income (i.e. country income normalized with respect to the arithmetic mean of world income). These observations can be plotted on a plane representing income mobility as shown in Figure 2. Contrary to the typical representation of such data, relative income at time *t* ($y \in S$) is represented on the *y*-axis and relative income at time $t + t$ ($x \in S$) on the *x*-axis. The axes have been inverted in order to graphically construct a transition probability matrix/kernel. Note the minimal variance of the observations about the diagonal and the varying density of the observations along the diagonal. The whole debate surrounding the application of the simple Markov chain model to panel data on income is how to properly extract the long-run distribution from this process that is characterized by these two very particular properties. This extraction requires the estimation of a transition probability matrix/kernel and so the problem of extracting the long-run distribution is translated into the problem of estimating the transition probability matrix/kernel corresponding to the income mobility process. In this section, a better understanding of the different approaches taken in solving this estimation problem, and the corresponding difficulties encountered, is achieved

[6] The 98 countries are: Algeria, Argentina, Australia, Austria, Barbados, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Burundi, Cameroon, Canada, Cape Verde, Chad, Chile, China, Colombia, Comoros, Congo, Republic of, Costa Rica, Cote d`Ivoire, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Finland, France, Gabon, Gambia, The, Ghana, Greece, Guatemala, Guinea, Guinea-Bissau, Honduras, Hong Kong, Iceland, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Korea, Republic of, Lesotho, Luxembourg, Madagascar, Malawi, Malaysia, Mali, Mauritius, Mexico, Morocco, Mozambique, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Romania, Rwanda, Senegal, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Syria, Taiwan, Tanzania, Thailand, Togo, Trinidad &Tobago, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zambia, and Zimbabwe.

by placing them in the context of Silverman's (1986) classic survey of non-parametric density estimation techniques.

Figure 2: Scatterplot of cross-county income mobility



In the Markov literature on cross-country convergence, the scatterplot of the observations is implicitly assumed to be a realization of the 'true' Markovian transition operator; consequently, the transition probability matrix/kernel estimation problem is implicitly treated as a standard non-parametric density estimation problem (Quah 1996b & 1997, Desdoigts 1996, Johnson 2000, Bulli 2001). In Sections 2.1 and 2.2, the discrete and continuous approaches to the estimation problem are explicitly presented as Silverman's histogram and kernel estimators. The treatment of the observations as a realization of the 'true' Markovian transition operator overlooks the widely cited fact that a Markov process retains the Markov property only for certain space and time discretizations (Kemeny and Snell 1960), and the empirical reality that whereas the observations take values belonging to the continuous space of relative incomes, they are collected only at discrete points in time. Therefore, in order for the observations to represent a realization of the 'true' Markovian transition operator, they need to be corrected for the most probably improper time discretization imposed by the annual frequency of the data.

Viewing the estimation problem from this new perspective, it appears that the efforts in the domain have been misdirected, relatively much attention in the literature having been accorded to the problem of space discretization (Aghevli and Mehran 1981, Davies and Shorrocks 1989, Magrini 1999, Bulli 2001), and absolutely no attention to the problem of time discretization. Indeed, because estimation of the Markovian transition operator is thought to be just a question of unwieldly non-parametrics, the number one priority has been to recover the simplicity inherent in the discrete approach without compromising the precision of the continuous approach (i.e. to develop a rigorous method of space-discretization); taking the implications of annually observed data into account, it appears that one step has been skipped in the race to simplicity and that the number one priority should be to extract a realization of the 'true' Markovian transition operator from the observations (i.e. to develop a rigorous method of time-discretization). In Section 2.3, Bulli's (2001) paper on regenerative discretization is reinterpreted to provide the required correction of the observations and thus generate a realization of the 'true' Markovian transition operator. Once this realization has been identified, estimation of the transition matrix/kernel and calculation of the corresponding ergodic distribution proves to be a very straightforward process.

2.1 Discrete approach to space discretization: the histogram estimator

The simplest and most popular solution to the transition operator estimation problem is maximum likelihood (ML) estimation of the transition probability matrix. There are a small number of states $\bar{i}$ $(i, j = 1,...,\bar{i})$ and transitions between these states are observed at regular intervals t for a finite length of time *T*. Let $N(t+t)$ be the $\bar{i}\times\bar{i}$ matrix of transitions observed at time $t+t$, where the *ij*-th element $n_{ij}(t+t) = N_{t+t}(y = j|x = i)$ represents the number of transitions from state *i* to state *j* observed at time $t+t$. Summing over the columns of the matrix of transitions observed at time $t+t$ provides the distribution of observations across the states at time *t*, denoted $n(t)$ where the *i*-th element $n_i(t) = \sum_{j=1}^{\bar{i}} n_{ij}(t+t)$ represents the number of observations in state *i* at time *t*. Summing over the rows of the matrix of transitions observed at time $t+t$ provides the distribution of observations across the states at time $t+t$, denoted $n(t+t)$ where the *j*-th element $n_j(t+t) = \sum_{i=1}^{\bar{i}} n_{ij}(t+t)$ represents the number of observations in state *j* at time $t+t$. Summing over the rows and columns of the matrices of transitions observed during each of the time periods provides the total number of observed transitions $n = \sum_{i=1}^{\bar{i}} \sum_{j=1}^{\bar{i}} \sum_{t=t}^{T} n_{ij}(t)$. Suppose that the observed transitions are generated by a time-homogenous Markov chain of order one according to the matrix of transition probabilities *P*, where the *ij*-th element $p_{ij} = P(y = j|x = i)$ represents the probability of transiting from state *i* to state *j*. Maximization of the log-likelihood function

$$L = \sum_{i=1}^{\bar{i}} \sum_{j=1}^{\bar{i}} \sum_{t=t}^{T} n_{ij}(t) \ln(p_{ij}) \quad \textit{such that} \quad \sum_{j=1}^{\bar{i}} p_{ij} = 1 \quad \forall\ i = 1,...,\bar{i} \tag{1}$$

yields the ML estimator of *P*, denoted $\hat{P}$, where the *ij*-th element is

$$\hat{p}_{ij} = \frac{\sum_{t=t}^{T} n_{ij}}{\sum_{j=1}^{\bar{i}} \sum_{t=t}^{T} n_{ij}}. \tag{2}$$

The model can be summarized by the following expression of the Markov property: $n(t+t) = n(t)\cdot P$. The ergodic distribution is then $n(\infty) = n(\infty)\cdot P$.

Five possible states are defined by discretizing the set of possible values of relative incomes into intervals at 0.2, 0.4, 0.8, and 2.0 (c.f. Section 3). The estimated transition matrix and the corresponding ergodic distribution are:

$$\hat{P}^{crisp} = \begin{bmatrix} 0.98 & 0.02 & 0 & 0 & 0 \\ 0.05 & 0.92 & 0.03 & 0 & 0 \\ 0 & 0.03 & 0.95 & 0.02 & 0 \\ 0 & 0 & 0.03 & 0.96 & 0.01 \\ 0 & 0 & 0 & 0.01 & 0.99 \end{bmatrix} \qquad (3)$$

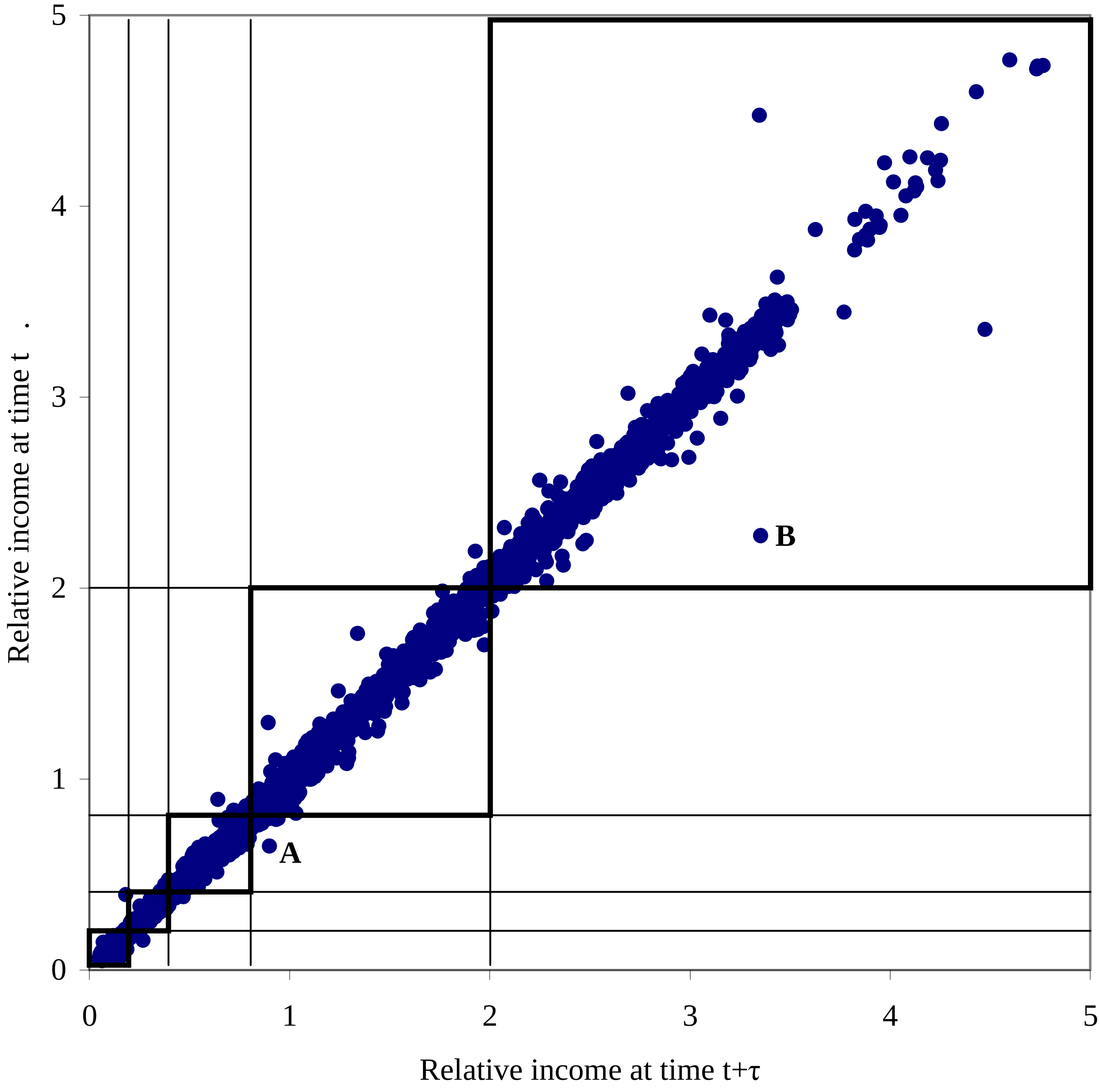
$$\hat{n}^{crisp}(\infty) = \begin{bmatrix} 0.38 & 0.15 & 0.15 & 0.12 & 0.20 \end{bmatrix}$$

As in Quah (1993a, 1993b, 1996a), the estimated ergodic distribution indicates an evolution towards a bipolar world of haves and have-nots.[7]

Let's take a closer look at the ML estimate of row *i* of the transition probability matrix: $p_i = \begin{bmatrix} \frac{n_{i1}}{n_i} & \ldots & \frac{n_{i\bar{i}}}{n_i} \end{bmatrix}$. This is nothing other than the histogram of observations made in state *i*, where class length (i.e. bin width) is normalized to one. Indeed, Equation 2 can be rewritten in a format similar to that used by Silverman (1986) on page 9 to define the histogram: $\hat{p}_i(x_{ij}) = \frac{1}{n_i \cdot 1}(n_{ij})$. Thus, the ML estimate of the transition matrix can be viewed as a stack of $\bar{i}$ such histograms. In view of comparing the different approaches to estimation of the transition matrix/kernel, it is useful to have an even more graphical understanding of what the discrete approach looks like. The ML estimate of the transition matrix can be obtained by superimposing a coarse (i.e. $\bar{i} \times \bar{i}$) grid upon the scatterplot of the observations (c.f. Figure 3), counting up the cell contents and putting them into relation with the row sums (c.f. Equation 3).

[7] The differences between the results presented in Quah (1993a, 1993b, 1996a) and those presented here are due to various factors. First, the data comes from a more recent version of the Penn World Tables, is composed of somewhat different countries, and is defined for a longer time period. Second, the discretization of the continuum of relative incomes needed to be marginally adjusted (c.f. Section 3).

Figure 3: Discrete Markov chain model approximation of cross-country income mobility



As established in Chapter 1, the ergodic distribution generated by the discrete estimate of the transition probability matrix proves to be so sensitive to marginal perturbations to the transition probabilities that a reasonably small confidence region for the estimated transition matrix generates an unreasonably large confidence region for the ergodic distribution (c.f. Equation 4).

$$\hat{P}^{crisp} \in \left\{ \begin{bmatrix} 0.974 & 0.009 & 0 & 0 & 0 \\ 0.031 & 0.906 & 0.018 & 0 & 0 \\ 0 & 0.019 & 0.932 & 0.012 & 0 \\ 0 & 0 & 0.016 & 0.948 & 0.005 \\ 0 & 0 & 0 & 0.001 & 0.988 \end{bmatrix}, \begin{bmatrix} 0.991 & 0.027 & 0 & 0 & 0 \\ 0.060 & 0.944 & 0.042 & 0 & 0 \\ 0 & 0.042 & 0.963 & 0.032 & 0 \\ 0 & 0 & 0.088 & 0.974 & 0.019 \\ 0 & 0 & 0 & 0.012 & 0.999 \end{bmatrix} \right\}$$

$$\hat{n}^{crisp}(\infty) \in \{[0.02 \quad 0.01 \quad 0.02 \quad 0.01 \quad 0.01], [0.80 \quad 0.34 \quad 0.38 \quad 0.39 \quad 0.87]\}$$

This extreme non-robustness of the estimated ergodic distribution renders this basic version of the discrete approach non-practicable. The only possibility of salvaging this simple approach lies in improving the precision of the discrete estimate of the transition matrix.

In view of improving upon the ML estimator of the transition probability matrix, let us take another look at Figure 3. If we look at the stack of histograms through Markovian lenses, the following gross inadequacy becomes evident. True immobility is represented by the line tracing the diagonal of the superimposed grid, and mobility increases with the distance from this diagonal line. Thus, in Figure 3, observation B represents more mobility than observation A represents, and yet this basic characteristic is not captured by the estimated model. Estimated immobility is represented by the transition probabilities located on the main diagonal of the transition matrix, but in Figure 3 mobile observation B falls into a main diagonal cell while immobile observation A falls into an off-diagonal cell. Two simple corrections to the estimation procedure outlined above help improve the correspondence achieved between true and estimated mobility. The most obvious correction would be to refine the grid superimposed upon the data plot. The other correction would be to 'nudge' the mobile observations that are trapped in off-diagonal cells just over the frontier into main diagonal cells. This 'nudging' can be achieved by filtering the time series of short-run noise (i.e. by requiring a transition to last a minimum number of periods for it to be included in the off-diagonal cell count, as in Chapter 1), or by smoothing the data over the time dimension (i.e. by replacing the time series by their moving medians, as in Section 5 of this chapter). Although these corrections improve the correspondence achieved between estimated and true mobility, by decreasing the off-diagonal cell counts, they actually exacerbate the problem of non-robustness of the estimated ergodic distribution.

If we now look at Figure 3 through non-parametric lenses, Silverman (1986) can be used to identify other inadequacies with the discrete approach to the transition operator estimation problem. The use of the histogram to accurately represent data is limited by the necessary choice of bin origin and bin width. This dependency of the estimate of the transition probability matrix upon class definition introduces just the sort of uncertainty in the accuracy of the discrete estimate that needs to be minimized in order to limit the confidence region of the ergodic distribution. The use of the more sophisticated non-parametric estimator, the kernel estimator, by centering bins upon the datapoints themselves, constitutes a well-known improvement upon the histogram.

2.2 Continuous approach to space discretization: the kernel estimator

The generally accepted solution to the transition operator estimation problem is non-parametric estimation of the transition probability kernel. There are an infinite number of states ($k, l \in S$), and transitions between these states are observed at regular intervals t for a finite length of time *T*. Let $N_{t+t}(x, y)$ be the joint density of transitions observed at time $t+t$, and let $n_t(x) = \int_0^\infty N(x, y)dy$ and $n_{t+t}(y) = \int_0^\infty N(x, y)dx$ be the distributions of observations across the states at times *t* and $t+t$ respectively (i.e. current and future marginal densities respectively). Normalizing the joint density with respect to the current marginal density provides the conditional density of transitions observed at time $t+t$, denoted $N_{t+t}(y|x)$. Suppose that the observed transitions are generated by a time-homogenous Markov process of order one according to the kernel of transition probabilities *P*. Non-parametric estimation of the conditional density of observed transitions yields the following bivariate kernel density estimator:

$$\hat{P}(y|x) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left\{\left[\frac{1}{h}(x-x_i) \quad \frac{1}{h}(y-y_i)\right]\right\}, \tag{5}$$

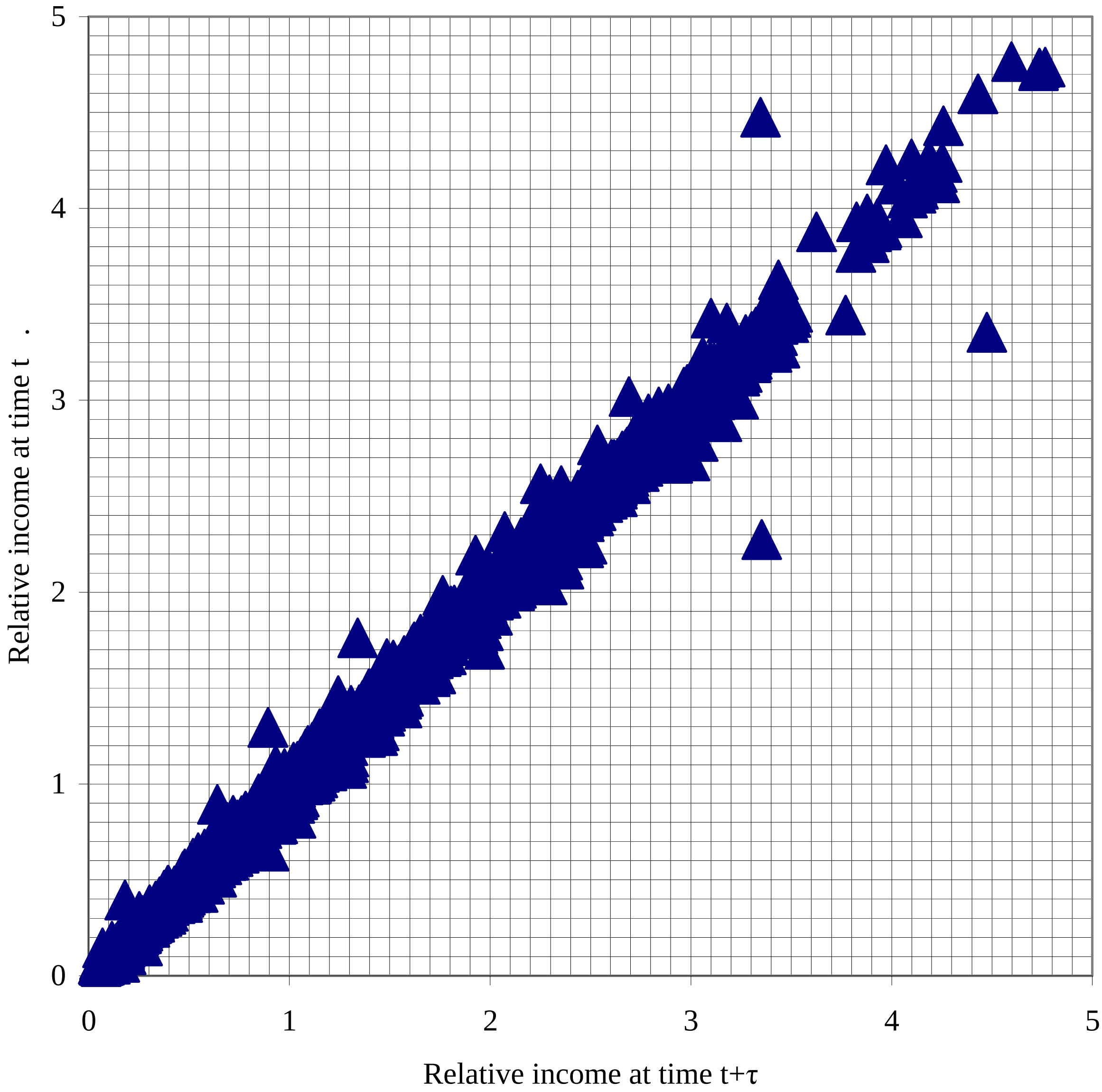
where $h = 1.2 * (\hat{s}_x^2 + \hat{s}_y^2) * n^{-1/6}$ is the rule-of –thumb window width, $\hat{s}_x^2$ and $\hat{s}_y^2$ are the estimated variances of *x* and *y* respectively, and *K* is the bivariate Epanechnikov kernel:

$$K = \begin{cases} \frac{2}{ph^2}\left\{h^2 - (x-x_i)^2 - (y-y_i)^2\right\} & if \ (x-x_i)^2 + (y-y_i)^2 < h^2 \\ 0 & otherwise \end{cases} \quad (6)$$

The model can be summarized by the following expression of the Markov property: $n_{t+t}(y) = \int_0^\infty P(y|x)n_t(x)dx$. The ergodic distribution is then $n_\infty(y) = \int_0^\infty P(y|x)n_\infty(x)dx$.

Let's take a closer look at the non-parametric estimation of the transition probability kernel in action. Whereas the theoretical definition of the discrete version of the Markov model is directly applicable, practical difficulties with the analytical evaluation of the continuous version of the Markov model oblige the researcher to proceed in a numerical fashion. This means evaluating the bivariate kernel density estimator at the coordinates of a very fine grid and then applying the discrete approach to this bigger and better matrix of observed transitions. Viewed in this light, a very concrete understanding of the improvements embodied in the continuous application of the Markov model is easy to attain. The first improvement is the use of the kernel function to smooth the observed transitions over the space dimension. The second improvement is the use of a large number of states to simulate continuity as best as possible. Both of these improvements correct for the poor correspondence between estimated and true mobility that is unavoidable in the discrete application of the Markov model. More technically, the numerical evaluation of the continuous estimate of the transition operator converges to the discrete estimate of the transition operator as the window width of the smoothing function converges to a point and as the gridunit used in the numerical evaluation of the continuous approach converges to the class width used in the discrete approach. More graphically, the observations are once again plotted upon the plane of income mobility, but this time the observations are represented by 'bumps' and not by points, then a grid is once again superimposed upon the data plot, but this time the grid is a very fine one and not a coarse one, and finally cell contents are once again counted up and put into relation with row sums (c.f. Figure 4).

Figure 4: Continuous Markov chain model approximation of cross-country income mobility



If the object is to estimate the density of observed transitions, then the very basic non-parametric approach outlined above must and easily can be improved upon. It must be improved upon because it is not designed to deal with the two characteristics of the mobility process noted in the introduction to this section, namely the minimal variance of the data about the diagonal and the varying density of the data along the diagonal. Using a constant bandwidth in such a situation means that the regions of high data density are oversmoothed and the regions of low data density are undersmoothed. In the controversy over the presence of twin peaks in the ergodic cross-country income distribution, it is essential that there are no spurious modes generated by undersmoothing. It can easily be improved upon because in Silverman (1986) there is considerable space dedicated to improvements upon the constant bandwidth chosen by rule-of-thumb, such as the adaptive kernel chosen by likelihood cross-validation. The fuzzification of income observations that I develop in Section 4.2 is a variant of the approach outlined above. The main difference lies in the fact that the bandwidth in the fuzzy approach does not vary with data density as above, but with data quality.

This perspective à la Silverman of the Markovian approach to the cross-country convergence debate redefines the whole estimation problem. Two separate issues emerge: space-discretization and time-discretization. In Sections 2.1 and 2.2, the time-discretization problem has been assumed away (c.f. "*Suppose that the observed transitions are generated by a time-homogenous Markov process of order one…*"), in order to deal with the space-discretization problem. In Section 2.3, the time-discretization problem will be discussed.

2.3 Regenerative sampling: time and space discretization

To my knowledge, time discretization has only been addressed in Bulli (2001).[8] The Silverman perspective that I take in this chapter imposes an alternative interpretation of the results presented in this paper. She compares three estimates of the ergodic distribution. The first estimate (i.e. 'the discrete limiting distribution obtained from a naïve discretization') is obtained via the simplest discrete analysis, via the superimposition of a coarse grid upon the scatterplot of the data (c.f. Figure 3). The second estimate (i.e. 'the continuous limiting distribution') is obtained via the simplest continuous analysis, via the superimposition of a fine grid upon the 'bump-plot' of the data in which all bumps are chosen to be identical (c.f. Figure 4). The third estimate (i.e. 'the discrete limiting distribution obtained from a regenerative discretization') is obtained via the superimposition of a coarse grid upon the scatterplot of simulated data.

Taking the second estimate to represent the truth, she rejects the first estimate and accepts the third estimate. In light of the previous discussion, taking the second estimate as

[8] Bulli (2001) presents regenerative sampling as a solution to the space discretization problem. More notably, it is also a solution to the time discretization problem.

the truth is already problematic (i.e. a constant bandwidth chosen by rule-of-thumb is just too crude), but overlooking this detail, other interesting information can be read from these results. On one hand, the rejection of the first estimate is not at all surprising; the naïve estimator is clearly an inadequate non-parametric estimator and the importance of smoothing over the space dimension is thus affirmed. On the other hand, the acceptance of the third estimator is a little more surprising.

What exactly is the difference between the second and third estimators? As already mentioned, the second estimate is obtained via the superimposition of a fine grid upon the 'bump-plot' of the data. The data for the third estimate is obtained by using this finely-gridded bump-plot of the data to generate observations in a Markovian way (more details are not pertinent). The third estimate is obtained via the superimposition of a coarse grid upon the 'bump-plot' of the selected data. So, the difference between the two estimates is the regenerative discretization, which means that affirming that both methods provide the same results, is affirming that the regenerative discretization does not make any difference!

The problem here is not that the regenerative discretization does not make any difference, but rather that the second estimator does not represent the truth. What is referred to as the transition probability kernel is actually a smoothed representation of the data from which the Markovian process still needs to be extracted. This implies defining a state-discretization AND a time-discretization for which the Markov property holds. Regenerative discretization does precisely this, and as such generates the data needed to construct the transition probability kernel.

My message can be boiled down to the following two points. First, the cross-country income convergence debate has focused on the second step of the problem, the identification of the Markov model (space-discretization), before having resolved the first step of the problem, the specification of the Markov model (time-discretization). Second, the quality of the space-discretization can easily be improved upon by taking into account the recommendations made in the classical literature on density estimation. In this chapter, I too am guilty of abstracting away from the time-discretization problem. My contribution is to improve upon the quality of the space-discretization of the Markov process. As already mentioned, my solution addresses not only the issue of fragility of the ergodic distribution, but also the issue of biasedness of the estimated transition probability matrix. This is the issue to which we now turn.

3. Biasedness of the estimated transition probability matrix

The results from the estimation of the simple Markov chain model are collected in Appendix A.1. As these results so strikingly illustrate, the information contained in the database, as summarized by the estimated transition probability matrix, is not sufficiently precise to generate any meaningful information whatsoever on the long-run tendencies. In other words, the sample size is too small to carry out significant statistical inference on the ergodic distribution. As shown in Chapter 1, the problem is not the total number of observations contained in the database, but rather the grossly unequal distribution of the observations amongst the cells of the observed transition matrix. This matrix informs us that the twin peaks displayed by the ergodic distribution are actually generated by only 23 of the 4214 total observations[9], that is by less than 1% of all observations. We are facing a problem of efficiency. As will be shown in this section, the application of a discrete model to continuous data poses another more fundamental problem to the precise estimation of the transition probability matrix.

The precision of the estimated transition probability matrix characterizing a process that is supposed to be Markovian is intimately related to the definition of classes. Indeed, a process is Markovian if there is one system of classification for which the Markov property holds, even if there is no other system of classification for which the Markov property holds (Kemeny and Snell, 1960). In other words, if the classes are not defined properly, then the model is misspecified.

The process of class definition can be decomposed into the initial division of the state space into homogenous groups and the subsequent choice in the level of aggregation. When data is discrete, there is a 'natural' division of the state space into groups, leaving the appropriate level of aggregation to be determined by the researcher only when necessary.[10]

[9] Twin peaks exist when $c > b$ and $m > n$ simultaneously; hence, the twin peaks here are generated by $(n_{21} - n_{12}) + (n_{45} - n_{54}) = (35-16) + (10-6) = 23$ observations (c.f. Section 4.2 in Chapter 1 and Appendix A.1 in this chapter).

[10] In the Markovian analysis of time changes in political attitudes in the United States, the nature of the subject

When data is continuous, however, there is no such 'natural' division of the state space and the whole process of class definition becomes somewhat arbitrary. Two criteria commonly employed in the literature to discretize the continuous state space include equal log length (Champernowne 1953, Shorrocks 1976, Aebi, Neusser and Steiner 2001) and equal number of observations (Quah 1993a 1993b 1996a, Proudman and Redding 1998, Proudman et. al. 1998). Two other criteria proposed in the literature include minimization of the area between the Lorenz curves corresponding to the continuous and discretized datasets (Aghevli and Mehran 1981, Davies and Shorrocks 1989) and minimization of errors generated by the use of a common discretization grid for the non-parametric representation of the data distributions in the first and last years of the dataset (Magrini 1998). The problem with all of these methods of discretization is that they completely ignore the Markovity of the problem, concentrating rather on the non-parametricity of the problem. The choice in the level of aggregation is then usually determined by choosing the maximum number of classes given a minimum number of observations per class. The virtual impossibility of thus correctly identifying the system of classification for which the Markov property holds is a fact that has been long and widely accepted in the literature. Bulli (2001) proposes a method of discretization of the continuous state space that automatically generates a system of classification for which the Markov property holds, but as previously mentioned, this solution compromises the simplicity of the Markovian framework.

In this chapter, the fuzzification of the income class frontiers that is carried out in Section 4.1 circumvents in large part the whole problem of class definition. As such, the strategy employed in class definition is entirely motivated by concerns for statistical robustness. Classes are defined such that each class contains roughly the same number of observations. There are two related reasons for this. Firstly, given the limited number of observed transitions, they need to be shared amongst the classes. Ensuring the robust estimation of one of the off-diagonal transition probabilities by allocating disproportionate numbers of observations via class definition would come at the cost of the robust estimation of the other off-diagonal transition probabilities. Secondly, given the limited number of

under study leaves little scope for variation in class definition. Anderson (1954) defines the only three classes possible: 'Republican', 'Democratic' and 'Don't Know'. In the Markovian analysis of intergenerational occupational mobility, on the other hand, the researcher plays a more active role in class definition. Glass and Hall (1954) define the following seven classes: 'Professional and higher administrative', 'Managerial and executive', 'Higher grade supervisory and non-manual', 'Lower grade supervisory and non-manual', 'Skilled manual and routine non-manual', 'Semi-skilled manual', 'Unskilled manual'. Prais (1955) takes a more aggregated approach and defines the following three classes: 'Upper', 'Middle', 'Lower'.

observed transitions, and given that the shape of the ergodic distribution is defined by the ratios between the probabilities of such transitions, these transition probabilities need to be representative of proportionately similar numbers of observations. Twin peaks generated by the preponderant weight of a couple of observations in a class of few observations against the weight of more observations in a class of many observations is not as robust (because of the small sample size) as a peak generated by a straightforward imbalance between equally weighted probabilities.

Note that such a strategy of class definition generally produces classes of variable length, as in the context of the cross-country convergence application where the class frontiers generated in this manner are: 0, 0.2, 0.4, 0.8, 2, 5. In this paragraph, a slight digression is made to examine the implications of variable class length for interpretation of the ergodic distribution. The values taken by the elements of the ergodic distribution are implicitly interpreted as the ***heights*** of the bins of the histogram approximating the long-run distribution. This interpretation is only valid if the classes are of equal length because the values taken by the elements of the ergodic distribution actually represent the relative ***areas*** of the bins of this histogram. So, when classes are of variable length, the ergodic distribution needs to be scaled by these variable lengths in order to generate an accurate picture of the histogram approximating the long-run distribution. When this rescaling is carried out in the context of the cross-country convergence application, twin peaks give way to a much bleaker picture of the future in which relative poverty traps most countries.

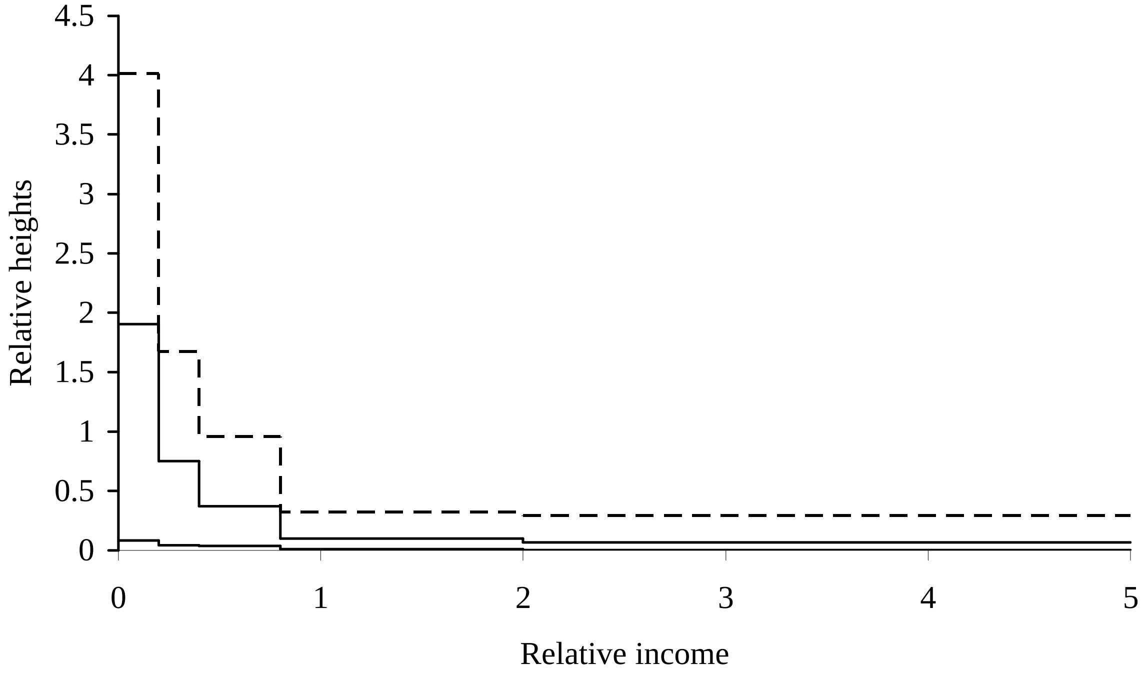
$$\begin{aligned} \hat{n}^{crisp}_{rescaled}(\infty) &= \begin{bmatrix} 1.90 & 0.75 & 0.37 & 0.10 & 0.07 \end{bmatrix} \\ \hat{n}^{crisp}_{rescaled}(\infty) &\in \left\{ \begin{bmatrix} 0.08 & 0.04 & 0.04 & 0.01 & 0.00 \end{bmatrix}, \begin{bmatrix} 4.01 & 1.68 & 0.96 & 0.32 & 0.29 \end{bmatrix} \right\} \end{aligned} \tag{7}$$

This corrected picture of the histogram approximating the long-run distribution (c.f. Figure 5) corresponds to the picture resulting from Bulli’s (2001) continuous and discrete Markovian analyses of the evolution of the cross-country income distribution.[11]

[11] There is a contradiction between the results presented in Table 2 and Figure 5 of Bulli (2001). The numerical results corroborate the poverty trap story, while the graphical results display a huge left-hand peak and a tiny right-hand peak.

Figure 5: Histogram of long-run tendencies

(solid line: rescaled estimated crisp ergodic distribution; dashed lines: upper and lower bounds defining 95% confidence region)

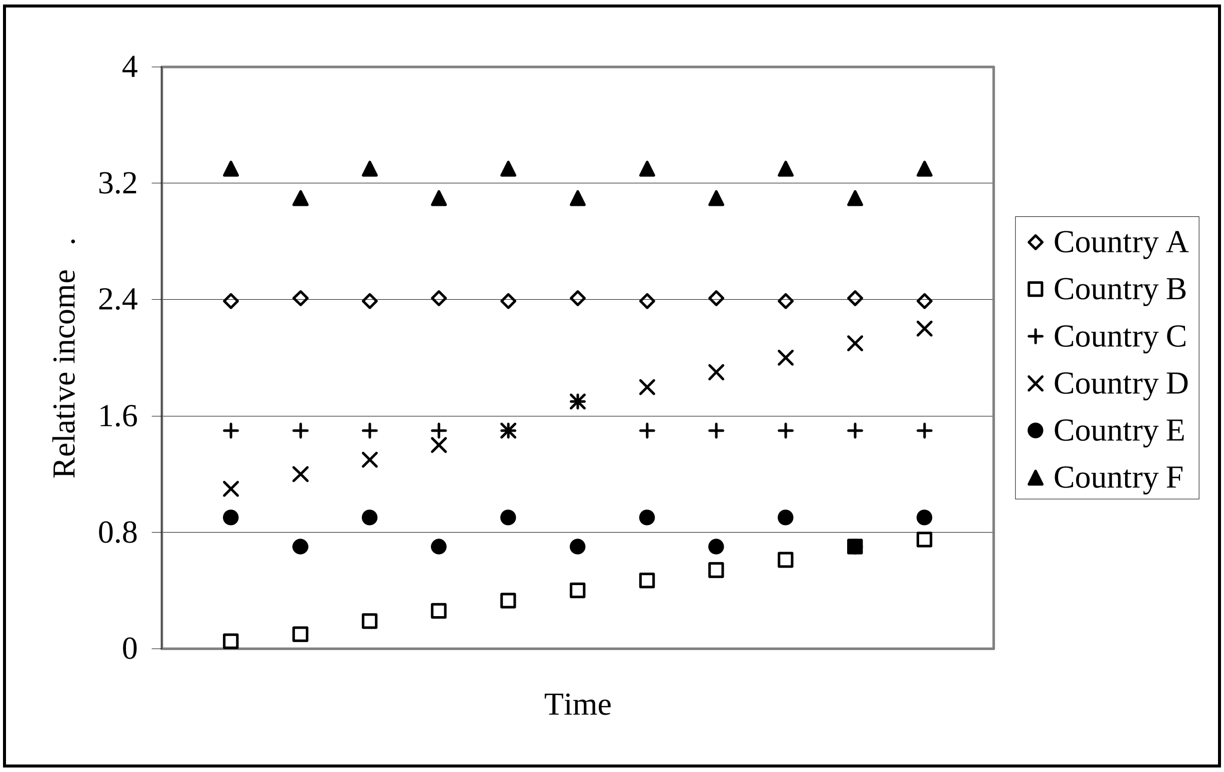


In sum, precise estimation of the transition probability matrix relies on the identification of a system of classification for which the Markov property holds. In the remainder of this section, it is shown that when data is continuous *any* definition of classes (i.e. including the one for which the Markov property holds) introduces a bias into the estimated transition probability matrix (Section 3.1), a problem that is exacerbated by the presence of data inaccuracies in general (Section 3.2.1) and differential data inaccuracies in particular (Section 3.2.2).

3.1 Discrete models and continuous data

In the simple Markov chain model, transitions are supposed to represent mobility. When classes are 'well-defined', transitions do represent mobility. By 'well-defined', I mean that observations within the same classes are relatively homogeneous and that observations within different classes are relatively heterogeneous. When data is discrete, the 'natural' division of the state space generates 'well-defined' classes, and mobility is easily identified. When data is continuous, there is no such 'natural' division of the state space, classes are not 'well-defined' and mobility is not as easily identified. The following two examples illustrate the implications of this ill-definition of classes for the precision of the estimated transition probability matrix.

Figure 6: Transitions resulting from scenarios presented in Examples 1 to 4



Example 1: Immobile process A is represented by mobile transition matrix P_A^c. Suppose that the evolution of country A's income closely follows the evolution of the world's average income, regularly alternating between periods of marginally less than average world growth and periods of marginally higher than average world growth. In other words, suppose that country A's income, as measured relative to the world average, remains essentially constant over time, oscillating between the values of 2.39 and 2.41. Figure 6 plots this surprisingly stable path over time through the distribution of income across countries. If the evolution of country A's relative income is modeled as a Markov chain, and if the underlying continuous

state space is discretized such that the relative income defining the frontier between classes 3 and 4 takes the value of 2.40, then the estimated transition probability matrix presents maximum mobility:

$$N_A^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad P_A^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad n_A^c(\infty)' = \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0.5 \\ 0 \end{bmatrix}. \tag{8}$$

In this particular case where the country's relative income marginally fluctuates about the level of relative income defining an income frontier, the simple Markov chain model provides a grossly misleading interpretation of the data. The problem lies in the continuity of the state-space; such problems do not arise when the state-space is discrete. Whereas there is no real difference between values of 2.39 and 2.41 for relative income when the object of study is mobility within the distribution of income across countries, there is a clear difference between baker and butcher (even if these professions could be considered to be very similar) when the object of study is mobility within the distribution of professions across the active population.

<u>Example 2</u>: Mobile process B is represented by immobile transition matrix P_B^c. Suppose that growth in country B's income consistently exceeds growth in the world's average income, starting the observation period at the bottom of the cross-country income distribution, and finishing the observation period at the top of the first class. Figure 6 plots this steady climb over time through the poorest class of the distribution of income across countries. If the evolution of country B's relative income is modeled as a Markov chain, and if the underlying continuous state space is discretized such that the relative income defining the frontier between classes 1 and 2 takes the value of 0.8, then the estimated transition probability matrix presents maximum immobility:

$$N_B^c = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad P_B^c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad n_B^c(\infty)' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{9}$$

In this particular case where the country's relative income grows within the limits defined by the frontiers of one single income class, the simple Markov chain model provides a grossly misleading interpretation of the data. Once again, the problem lies in the continuity of the state-space; such problems do not arise when the state-space is discrete. Whereas the intra-class growth in values of relative income from 0.05 to 0.75 is significant when the object of study is mobility within the distribution of income across countries, the intra-class mobility generated by the passage from cardiologist to hematologist is not significant at all when the object of study is mobility within the distribution of professions (not specialties) across the active population.

3.2 Data inaccuracies

In the previous section, data inaccuracy was ignored. Given the huge sensitivity of the elements of the ergodic distribution to marginal perturbations to the transition probabilities, however, all potential sources of imprecision in the estimated transition probability matrix must be examined. In this section, two aspects of data inaccuracy are discussed, namely its existence and its variation across observations. It is shown how the simple Markov chain model is particularly sensitive to the existence of data inaccuracy, especially when it affects different observations to different degrees.

3.2.1 Short-run noise vs. short-run dynamics

When data is continuous, the arbitrariness of the division of the state-space leads to ill-defined classes and falsely generated transitions. When data contains inaccuracies, these can also contribute to generating mobility-unrelated transitions. This problem is particularly acute when data is continuous. If data inaccuracy is composed of encoding and measurement errors, it seems reasonable to assume that whereas both continuous and discrete data are similarly affected by encoding errors, the nature of continuous data makes it more prone to measurement error than discrete data. As illustrated below in Example 3, the lack of distinction between genuine transitions generated by short-run dynamics, and false transitions generated by short-run noise exacerbates the bias already imposed upon the estimated transition probability matrix by the use of a discrete model with continuous data.

<u>Example 3</u>: *Short-run noise in process C and short-run dynamics in process D generate identical entries in the observed transition matrix.* Consider a world of uncertainty in which data may contain errors. Suppose that the estimate of country C's relative income remains essentially constant over time, deviating only once from its second class value of 1.5 to take on the third class value of 1.7. Suppose that the estimate of country D's relative income grows consistently, starting the observation period in the middle of the second class of the cross-country income distribution, and finishing the observation period in the middle of the third class. Figure 6 plots the paths of these two countries through the distribution of income across countries. If these paths are modeled as Markov chains, then the transition from class two to three that is shared by these very different countries during the fifth unit of observation generates identical observed transition matrices:

$$N_C^c(t=5) = N_D^c(t=5) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{10}$$

In our world of uncertainty, the robustness of the transition from class two to three is convincing when it occurs in the context of country D's relative growth, but questionable when it occurs in the context of country C's relative stagnation. In this realistic setting where short-run noise and short-run dynamics co-exist, the simple Markov chain model can provide a misleading interpretation of the data. The problem lies in the crippling effect of the presence of inaccuracies in the data upon the power of the Markov hypothesis to accurately represent the data. More specifically, the short-sightedness of the simple Markov chain model prevents it from using all of the information available to distinguish between identical transitions occurring in very different contexts.[12]

[12] Just to be absolutely clear, it is not the Markovian characterization of the data that is problematic (i.e. this is not a question of model misspecification), but rather the detection of the true data underlying the error inaccuracies via the Markov hypothesis that poses a problem (i.e. this is a question of model identification).

3.2.2 Data inaccuracy differentials

Precise estimation of the transition probability matrix requires distinguishing between short-run dynamics and short-run noise. Even when the context provides clear identification of the nature of the variation, the simple Markov chain model is incapable of making such a distinction. As illustrated below in Example 4, this problem is further complicated by the existence of differences in the extent to which different observations are affected by data inaccuracy.

<u>Example 4</u>*: Very imprecise series of data generated by process E and very precise series of data generated by process F contribute equally to generating mobility in the estimated transition probability matrix.* Consider a world of relative uncertainty in which data may contain different degrees of error. Suppose that poor country E's very imprecise estimate of relative income and rich country F's very precise estimate of relative income display identical fluctuations about the frontier between classes one and two and the frontier between classes four and five, respectively. Figure 6 plots the paths of these two countries through the distribution of income across countries. If these paths are modeled as Markov chains, then the transitions generated by these two series of very different quality data contribute equally to generating mobility in the estimated transition probability matrix:

$$N_E^c = \begin{bmatrix} 0 & 5 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad P_E^c = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad n_E^c(\infty)' = \begin{bmatrix} 0.5 \\ 0.5 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{11}$$

$$N_F^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 5 & 0 \end{bmatrix} \qquad P_F^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad n_F^c(\infty)' = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.5 \\ 0.5 \end{bmatrix}. \tag{12}$$

In our world of relative uncertainty, the robustness of the transitions is convincing when the underlying data is very precise, but questionable when it is very imprecise; whereas the fluctuations in country F's relative income could possibly reflect some sort of short-run dynamics, the fluctuations in country E's relative income most probably reflect some sort of short-run noise. In this very realistic setting where short-run noise and short-run dynamics co-exist in different proportions depending upon the quality of the data, the simple Markov chain model can provide a misleading interpretation of the data. The problem here is not one that is generated specifically by the application of the simple Markov chain model; the problem posed by differential data quality for inference is one that adversely affects many different models (Dawson, DeJuan, Seater, and Stephenson 2001).

As shown in this section, the continuity and the (differential) inaccuracy of the data pose fundamental problems for the precise estimation of the transition probability matrix. In Section 4, it is shown how very basic notions of fuzzy logic can be used to address most of these problems, rendering the simple Markov chain model amenable to continuous data of differential inaccuracy. In Section 5, it is shown how selective filtering can be used to address the rest of these problems, rendering the data amenable to the simple Markov chain model.

4. Fuzzification of the simple Markov chain model

What is fuzzy logic? In a nutshell, it is the rejection of a binary representation of the world in favour of a more realistic model. Things do not have to be black or white (i.e. crisp); they can be different shades of grey (i.e. fuzzy). By the same token, observed transitions should not have to take either zero or one as values; they should be able to take an intermediate value that is representative of the true degree of mobility characterizing the transition. Presumably, this added flexibility should help resolve the problems illustrated in Examples 1, 2 and 4.[13]

This argument can also be expressed in a more formal manner. In the application of the simple Markov chain model to panel data on income, the estimated transition probability matrix is used to extract information concerning the mobility of countries within the distribution of incomes from the dataset, information that is camouflaged by noise. In other

[13] The problems illustrated in Example 3 will be addressed in Section 5.

words, true mobility $m_{ij}(t)$ is observed with error $e(t)$, as described in the following expression.

$$n_{ij}(t) = m_{ij}(t) + e(t) \tag{13}$$

In empirical applications of the simple Markov chain model, the noise is overlooked, and the following assumptions are implicitly made.

$$\begin{aligned}
&P^c(m_{ij}(t)=0 \mid n_{ij}(t)=0)=1\\
&P^c(m_{ij}(t)=0 \mid n_{ij}(t)=1)=0\\
&P^c(m_{ij}(t)=1 \mid n_{ij}(t)=1)=0\\
&P^c(m_{ij}(t)=1 \mid n_{ij}(t)=1)=1
\end{aligned} \tag{14}$$

By explicitly taking the noise into account, it is possible to abandon this binary representation of the data in favour of a more detailed description of the data, as described in the following expressions:

$$\begin{aligned}
&P^f(m_{ij}(t)=0 \mid n_{ij}(t)=0)=1-f(e(t))\\
&P^f(m_{ij}(t)=0 \mid n_{ij}(t)=1)=0+f(e(t))\\
&P^f(m_{ij}(t)=1 \mid n_{ij}(t)=0)=0+f(e(t))\\
&P^f(m_{ij}(t)=1 \mid n_{ij}(t)=1)=1-f(e(t))
\end{aligned} \tag{15}$$

where f is a function of the error term. Here, the error term is composed of three elements, namely model misspecification, short-run noise and differential data inaccuracy (c.f. Section 3). In this section, fuzzification is used as a means of correcting the observed transition matrix for the first and the third of these elements. In the next section, selective filtration is used as a means of correcting the observed transition matrix for the second element of the error term. These corrections greatly improve the correspondence between the underlying data and the estimated transition probability matrix.

How can fuzzy logic be implemented in the context of the cross-country convergence debate? Translating the whole class definition issue (c.f. Section 3.1) into fuzzy terms, it is the crispness of the frontiers chosen to delineate the fundamentally fuzzy classes that poses problems. Fuzzifying income class frontiers solves these problems. From a philosophical point of view, this fuzzification captures the inherent vagueness of the income group notion (i.e. classes become ‘well-defined’). From a modelling point of view, this fuzzification replaces some of the discreteness in the simple Markov chain model with continuity, by taking into account the mobility occurring throughout the distribution and not just the mobility occurring at the frontiers. From a statistical point of view, the fuzzification increases the robustness of the estimated ergodic distribution by increasing the numbers of observations generating the transitions. Fuzzification of the income class frontiers is presented below in Section 4.1. In the first subsection, the intuition is presented via examples, and in the second subsection, the optimal degree of fuzzification to be carried out is discussed.

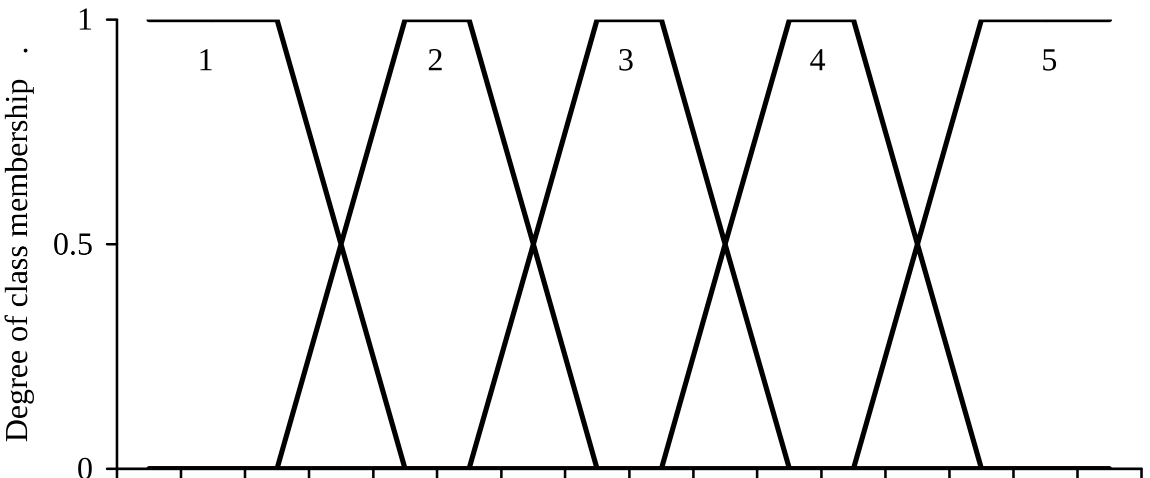
Translating the whole differential data inaccuracy issue (c.f. Section 3.2.2) into fuzzy terms, it is the uniform crispness of the differentially accurate observations that poses problems. Fuzzifying income observations solves these problems. From a modelling point of view, this fuzzification filters out short-run noise in a differential manner by directly incorporating the degree of inaccuracy inherent to the different observations into the method of accounting transitions. From a statistical point of view, once again the fuzzification increases the robustness of the estimated ergodic distribution by increasing the numbers of observations generating the transitions. Fuzzification of the income class frontiers is presented below in Section 4.2. In the first subsection, the intuition is presented via examples, and in the second subsection, the optimal degree of fuzzification to be carried out is discussed.

4.1 Fuzzification of the income class frontiers

4.1.1 Intuition and examples

Fuzzification of the class frontiers is achieved by defining the classes as overlapping trapezoidal distributions intersecting at the points demarcating the crisp class frontiers (c.f. top half of Figure 7), instead of defining the classes as uniform distributions over the intervals demarcated by the crisp class frontiers. This implies defining a distribution over classes per country per unit of time (c.f. bottom half of Figure 7), instead of defining a single class per country per unit of time; observations of classes become observations of distributions over classes. For example, an observation of a country with a relative income of 0.8 will no longer be allocated exclusively to class 2, but rather equally divided between class 1 and class 2.

Figure 7: Fuzzification of the income class frontiers



	0.0	0.3	0.5	0.8	1.1	1.3	1.6	1.9	2.1	2.4	2.7	2.9	3.2	3.5	3.7	4.0
1	1	1	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0.5	1	1	0.5	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0.5	1	1	0.5	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0.5	1	1	0.5	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0.5	1	1	1

Relative income

In this manner, countries are no longer forced into a strictly ‘poor’, ‘average’ or ‘rich’ classification; countries can be described as mixes between neighbouring classifications (for example, ‘slightly poor, mostly average’). This conversion of the set of possible observations per class from whole numbers to rational numbers automatically generates rational values for the observed transitions as well. As illustrated below in the continuations of Examples 1 and 2, estimating the transition probability matrix on the basis of this more detailed tabulation of transitions improves the correspondence between the mobility displayed by the time series of the data and the mobility displayed by the estimated transition probability matrix.

<u>Example 1 continued</u>: Immobile process A is represented by correspondingly immobile transition matrix P_A^f *.* The observations made of country A's path through the distribution of income across countries over time in the context of the crisp and fuzzy Markov chain models are collected below in A^c and A^f respectively:

$$A^c \equiv \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\} \quad A^f \equiv \left\{ \begin{bmatrix} 0 \\ 0 \\ 0.54 \\ 0.46 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0.46 \\ 0.54 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0.54 \\ 0.46 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ 0.54 \\ 0.46 \\ 0 \end{bmatrix} \right\}. \tag{16}$$

Whereas both A^c and A^f record the existence of the fluctuations in country A's path, only A^f supplies information on the nature (i.e. magnitude) of these fluctuations. When A^f is used to tabulate the numbers of observed transitions, and thus estimate the transition probability matrix and calculate the ergodic distribution, the following results are obtained:

$$N_A^f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.81 & 0.02 & 0 \\ 0 & 0 & 0.02 & 4.81 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad P_A^f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.96 & 0.04 & 0 \\ 0 & 0 & 0.04 & 0.96 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad n_A^f(\infty)' = \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0.5 \\ 0 \end{bmatrix}. \tag{17}$$

Note that fuzzification of the income class frontiers corrects for the misleading mobility biasing P_A^c, while retaining the very reasonable estimate of ergodic distribution.

<u>Example 2 continued</u>: Mobile process B is represented by correspondingly mobile transition matrix P_B^f . The observations made of country B's path through the distribution of income across countries over time in the context of the crisp and fuzzy Markov chain models are collected below in B^c and B^f respectively:

$$B^c \equiv \left\{ \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \ldots, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix} \right\}$$

$$B^f \equiv \left\{ \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0.99\\0.01\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0.86\\0.14\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0.69\\0.31\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0.59\\0.41\\0\\0\\0\end{bmatrix} \right\} \tag{18}$$

Whereas B^c tells the overly simplified story of inter-class immobility, B^f tells the more detailed story of intra-class growth. When B^f is used to tabulate the numbers of observed transitions, and thus estimate the transition probability matrix and calculate the ergodic distribution, the following results are obtained:

$$N_B^f = \begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&0&9.13&0.41&0\\0&0&0&0.47&0\\0&0&0&0&0\end{bmatrix}, \quad P_B^f = \begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&0&0.96&0.04&0\\0&0&0&1&0\\0&0&0&0&0\end{bmatrix}, \quad n_B^f(\infty)' = \begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}. \tag{19}$$

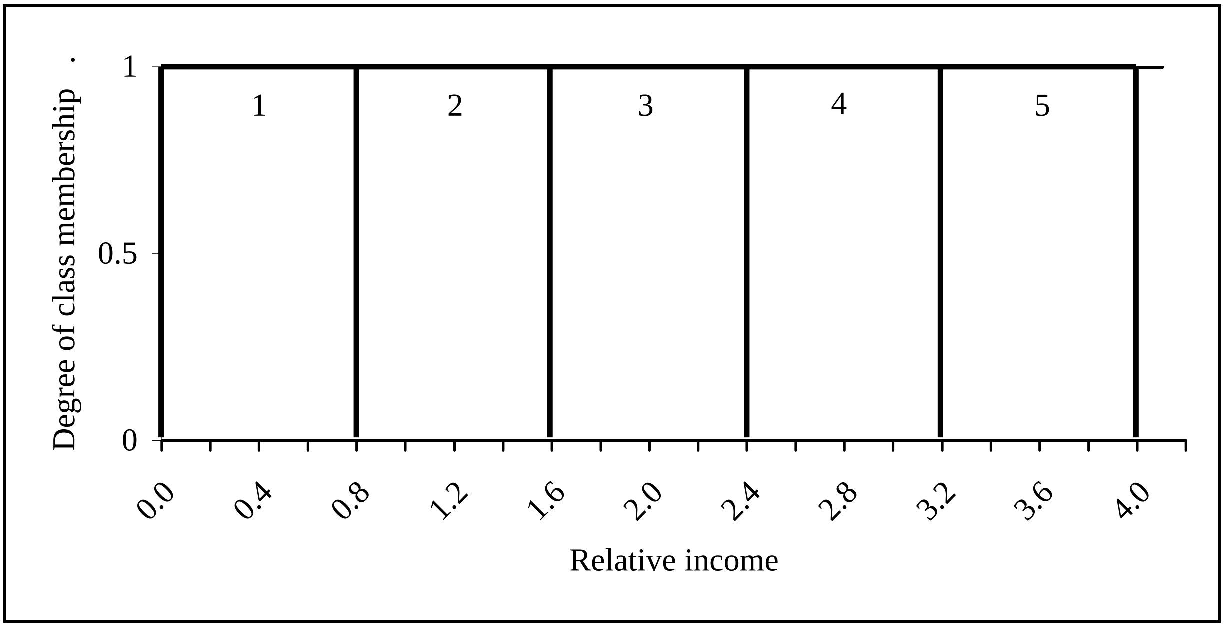
Note that fuzzification of the income class frontiers corrects for the misleading immobility biasing both P_B^c and $n_B^c(\infty)$.

4.1.2 How much fuzzification?

We have seen that fuzzification of the income class frontiers can be seen as the replacement of the crisp class frontiers by the intersections of trapezoidal distributions that are centered upon class midpoints. How should these trapezoids be chosen? Our only technical constraint is that the degrees of class membership must sum to one for each value of the continuum of relative incomes. Indeed, these values are probabilities and probabilities must sum to one. In Figure 7 this can be understood as requiring that the vertical sum of the distributions adds up to one. To this technical constraint, we can add a couple of common sense considerations that serve only to simplify the calculations. First, all class frontiers can be considered to be equally fuzzy. There is no reason to consider that the distinction between upper middle class and upper class is more clearly defined than that between lower class and lower middle class. Second, only the class frontiers between neighbouring classes are fuzzified, that is only adjacent trapezoidal distributions overlap. Defining an observation as a mix of three classes does not provide more information on its position within the continuum of relative incomes than defining the observation as a mix of two classes. The technical constraint and these two considerations imply that overlapping portions of the lateral sides of the trapezoids must have the same slope in absolute value and must intersect at the class frontiers at a class membership level of 0.5, and that the top and bottom sides of the trapezoids must sum to the width of the corresponding classes plus half the width of each of the neighbouring classes. Note that when classes are not of equal length, the lateral sides of the "trapezoids" become kinked (c.f. Figure 12). The degree of fuzzification can then be defined in terms of the width of the top side of the trapezoid relative to the width of the class.

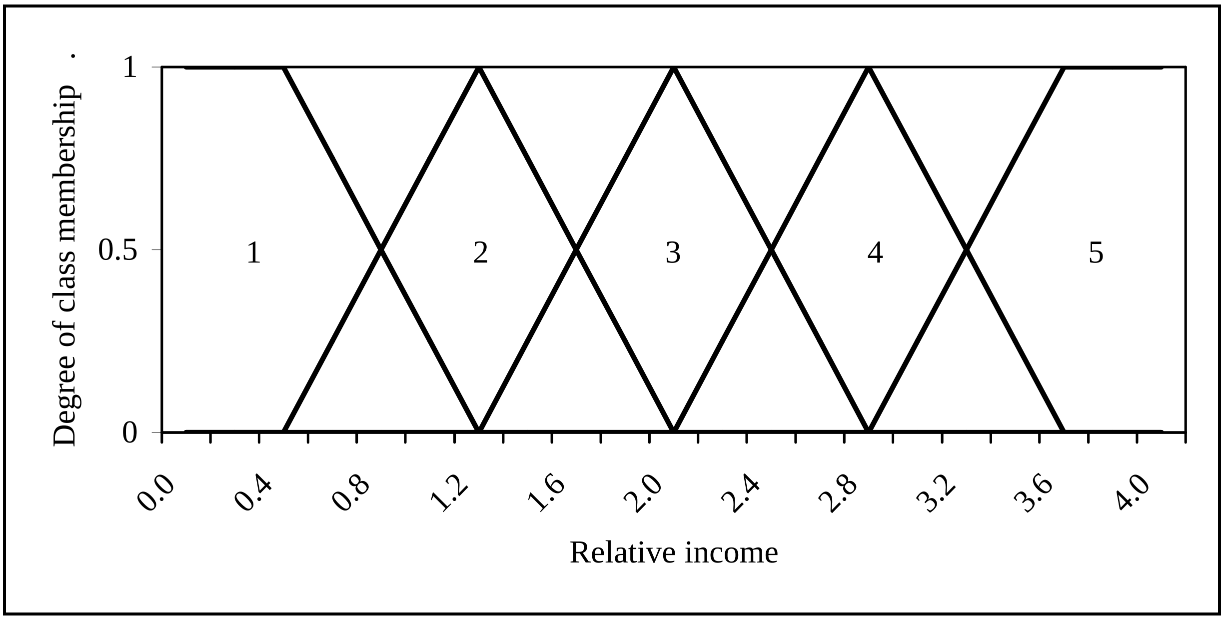
Minimum fuzzification occurs when the width of the top side of the trapezoid is equal to the width of the class (c.f. Figure 8).

Figure 8: Minimum fuzzification of the income class frontiers



In this case, the trapezoid is actually a rectangle, and the classes are defined as uniform distributions over the intervals demarcated by the crisp class frontiers. Observations can only belong to one class at a time. Maximum fuzzification occurs when the width of the top side of the trapezoid is a point (c.f. Figure 9).

Figure 9: Maximum fuzzification of the income class frontiers



In this case, the trapezoid is actually a triangle, and the classes are defined as triangular distributions over the intervals demarcated by the crisp class frontiers. Observations always belong to a mix of two neighbouring classes, except for those occurring exactly at the class midpoints. The examples in the previous section present an intermediate level of fuzzification (c.f. Figure 7) with observations occurring in the neighbourhood of the class midpoints recorded as only one class, and the others recorded as mixes of two classes.

In order to choose the most appropriate level of fuzzification, it helps to think about why we are carrying out this fuzzification. Fuzzification of the income class frontiers is a quick fix to some of the problems that arise when the simple Markov chain model is applied to continuous data (c.f. countries A and B in Examples 1 and 2). On the one hand we have continuous data which is rich in information but complex in analysis, and on the other hand we have the discrete Markov chain model which is simple in analysis but poor in information. This fuzzification allows one to apply the discrete Markov chain model to continuous data, without sacrificing information and compromising the validity of the results. It allows observations not only to be allocated to a class (as usual in the discretization of continuous data), but to be positioned within that class as well. How exactly the observations are positioned within the classes depends upon the degree of the fuzzification that is chosen. Minimum fuzzification means that once the observations have been classified, all additional information is lost. The intermediate fuzzification that is carried out in the previous section retains additional information on the positions within the classes for those observations that occur within the neighbourhood of the class frontiers, but not for those that occur within the neighbourhood of the class midpoints. Maximum fuzzification means that all the information contained within the continuous data is retained after the discretization. This is exactly what we are looking for, a simple approach that does not sacrifice the wealth of information provided by the continuity of the data.

So, is there really no reason for choosing less than maximum fuzzification of the income class frontiers? When applying the discrete Markov chain model to continuous data, the answer is yes. Contrary to the fuzzification of the income observations that is discussed in the next section, the choice of the degree of fuzzification of the income class frontiers cannot be thought of in the same terms as the choice of bandwidth of the kernel in non parametric density estimation. The key difference is that the choice of window width determines how much the *observations* are smoothed, whereas the choice of the degree of fuzzification determines how much the *model* is smoothed. Fuzzifying the income class frontiers is like rendering the discrete Markov chain model continuous while retaining the simplicity inherent in the discrete approach. So if the data is continuous, then the model must be adapted accordingly and completely. Partial fuzzification makes no sense. In Section 6.1, estimation of the simple Markov chain model with maximum fuzzification of the income class frontiers is carried out.

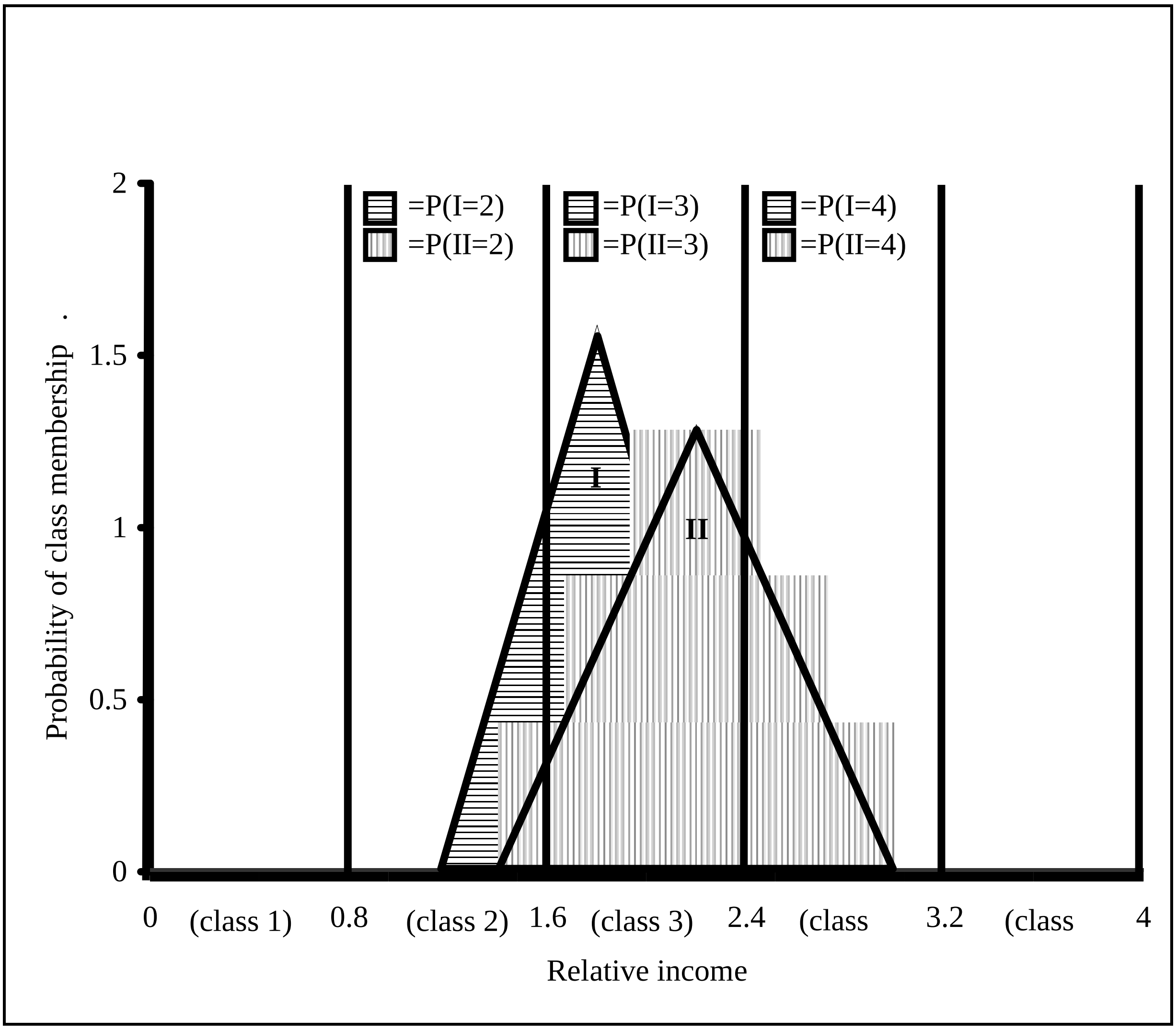
4.2 Fuzzification of the income observations

4.2.1 Intuition and examples

Fuzzification of the income observations is achieved by transforming the points observed on the continuum of relative incomes into triangular distributions centered on these points and defined over intervals that are directly proportional to the precision of the observations. In this manner, identical incomes observed for countries with different quality data are represented differently; observations from countries with relatively precise data are represented by relatively small-based triangles, and observations from countries with relatively imprecise data are represented by relatively large-based triangles. In order to integrate this new representation of the data into the simple Markov chain model, a new method of accounting transitions is required.

The construction of the matrix of observed transitions is perhaps best explained with the aid of a concrete example. Consider a country's grade 'D' relative income at two successive points in time. Grade "D" means that the confidence intervals for the observations can be constructed by taking the observation and subtracting off 30-40% of its value for the lower bound, and adding on 30-40% of its value for the upper bound. A value of 35% is used to carry out the calculations. The first observation of relative income is 1.8, so the corresponding confidence interval is [1.17,2.43]. This observation can be fuzzified by replacing the point observation at 1.8 by a triangle with a base defined over [1.17,2.43], an apex positioned at 1.8 and an area equal to 1. This gives the distribution that is labelled "I" in Figure 10. The second observation of relative income is 2.2, so the corresponding confidence interval is [1.43,2.97]. Fuzzifying the observation gives the distribution that is labelled "II" in Figure 10.

Figure 10: Fuzzification of the income observations



Note that the observations lie within classes 2, 3 and 4. This means that the transition from I to II generates entries in the observed transitions matrix for those elements belonging to the three-by-three sub-matrix containing the observed transitions for classes 2, 3 and 4. The area under each of the distributions is divided into the parts occupying the different classes. Each part represents the probability that the observation belongs to that class; in this example, we have: $P(I \in 2) = 0.233$, $P(I \in 3) = 0.766$, $P(I \in 4) = 0.001$, $P(II \in 2) = 0.024$, $P(II \in 3) = 0.702$ and $P(II \in 4) = 0.274$. These probabilities can be used as shown below to calculate the probabilities associated with all of the possible transitions.

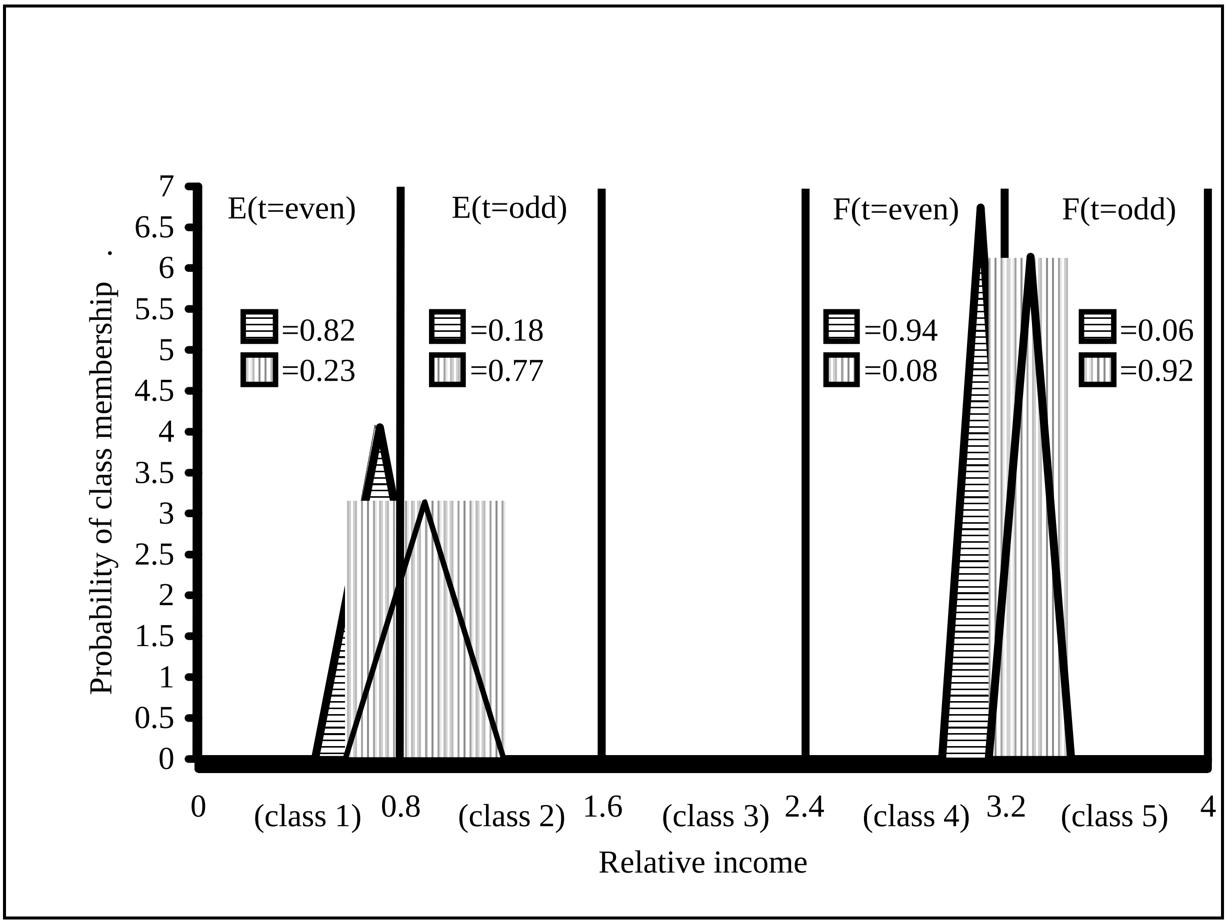
$$N^f(t) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & P(I \in 2)P(II \in 2) & P(I \in 2)P(II \in 3) & P(I \in 2)P(II \in 4) & 0 \\ 0 & P(I \in 3)P(II \in 2) & P(I \in 3)P(II \in 3) & P(I \in 3)P(II \in 4) & 0 \\ 0 & P(I \in 4)P(II \in 2) & P(I \in 4)P(II \in 3) & P(I \in 4)P(II \in 4) & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.006 & 0.163 & 0.064 & 0 \\ 0 & 0.019 & 0.537 & 0.210 & 0 \\ 0 & 0.000 & 0.001 & 0.000 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (20)$$

Carrying out these calculations for each pair of successive observations and summing over all of the resulting matrices generates the matrix of observed transitions. Estimation of the transition probability matrix and the ergodic distribution can then be carried out in the usual manner.

In sum, this way of accounting transitions is based on the implicit assumption that the part of the data's variance that is representative of true mobility is a positive function of quality. In other words, this way of accounting transitions is equivalent to adjusting the variance of the data in a differential manner, or to filtering the data of the differential short-run noise. As illustrated below in the continuation of Example 4, estimating the transition probability matrix on the basis of this more detailed tabulation of transitions improves the correspondence between the quality of the data and the amount of mobility extracted from the data by the estimated transition probability matrix.

<u>Example 4 continued</u>. *Very imprecise series of data generated by process E and very precise series of data generated by process F contribute differentially to generating mobility in the estimated transition probability matrix.* Suppose that the quality grades given to the data coming from countries E and F are 'D' and 'A' respectively. Grade 'D' remains defined as before. Grade 'A' means that the data is observed with a precision of 5-10%. A value of 5% is used to carry out calculations. Figure 11 provides a picture of the fuzzification of the two series of data.

Figure 11: Fuzzification carried out in Example 4



Note the marked difference in the probabilities calculated for the two identically fluctuating series of different quality data. The results generated by the fuzzy Markov chain model are presented below.

$$N_E^f = \begin{bmatrix} 1.92 & 3.37 & 0 & 0 & 0 \\ 3.37 & 1.34 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} P_E^f = \begin{bmatrix} 0.36 & 0.64 & 0 & 0 & 0 \\ 0.71 & 0.29 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} n_E^f(\infty)' = \begin{bmatrix} 0.53 \\ 0.47 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (21)$$

$$N_F^f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.73 & 4.35 \\ 0 & 0 & 0 & 4.35 & 0.58 \end{bmatrix} P_F^f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.14 & 0.86 \\ 0 & 0 & 0 & 0.88 & 0.12 \end{bmatrix} n_F^f(\infty)' = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.51 \\ 0.49 \end{bmatrix} \qquad (22)$$

Note that the mobility extracted from the fluctuations of the grade ‘A’ data is superior to that extracted from the same fluctuations present in the grade ‘D’ data (i.e. the off-diagonal elements of P_E^f are bigger than the corresponding elements of P_F^f). Thus, fuzzification of the income observations cleans up part of the short-run noise biasing both P_E^c and P_F^c .

4.2.2 How much fuzzification?

We have seen that fuzzification of the income observations can be seen as the recording of the observations as triangular distributions rather than as points. How should these triangles be chosen? Replacing points by distributions is just smoothing over the space dimension. This brings us back to the issue of non parametric density estimation that is discussed in Section 2.2. Viewed in this light, the question of the degree of fuzzification can be thought of in terms of the choice of the kernel’s bandwidth. As already discussed, a constant bandwidth is not appropriate to the data under analysis because this leads to oversmoothing some parts of the density and undersmoothing other parts of the density. The adaptive kernel addresses this drawback by allowing bandwidth to vary with data density (Silverman 1986). Fuzzification of the income observations addresses this drawback by allowing bandwidth to vary with data quality. This latter approach presents two advantatges when compared to the former approach. First, fuzzification is based upon economic criteria rather than some statistical criteria related to error minimization. It provides a way of including additional information on the quality of the data in the estimation process. Second, whereas the adaptive kernel can be chosen according to many different methods and therefore requires that the most appropriate method be chosen, fuzzification provides a natural answer to the bandwidth question. The bases of the triangles should span the confidence intervals of the observations. These advantages of the fuzzification approach can also be its disadvantage in the sense that this approach requires that information on the data quality be available, which is often not the case.

Thinking about fuzzification of the income observations in terms of non parametric density estimation brings up another related question. Why triangular distributions? Other kernels such as the Epanechnikov or the Gaussian kernels could conceivably be used as well, but in the absence of any information on the distribution of the errors in the data, there is no particular reason to further complicate calculations. In Section 6.2, estimation of the simple Markov chain model with fuzzification of the income observations is carried out.

5. Selective filtration of short-run noise

In the previous section, fuzzification of the simple Markov chain model was shown to correct for two of the three elements biasing the estimated transition probability matrix, namely the model misspecification resulting from the application of a discrete model to continuous data and the cross-sectional variations in short-run noise generated by the differentials in data inaccuracy. What these two elements have in common is that they both affect the empirical application in a general fashion; the problems posed by the model misspecification concern the full lengths of all of the income class frontiers, and the problems posed by the differential data inaccuracy concern the full lengths of the time series belonging to the different countries. The third element biasing the estimated transition probability matrix, the presence of short-run noise, differs from the first two elements in that it affects the empirical application in a very specific fashion; the problems posed by the existence of short-run noise are only generated by certain observations. Whereas the general model re-specification provided by fuzzification addresses the fundamental characteristics defining the problems generated by the model misspecification and the differential data inaccuracy, a different approach is required to solve the problems related to the presence of short-run noise.

Fuzzification of income observations resolves the problem of distinguishing between identical fluctuations generated by different quality data. Here the problem is to distinguish between identical fluctuations generated by same quality data, but having occurred in different contexts. The question is how to distinguish between short-run noise and short-run dynamics over time?

A quick and easy solution to this problem is to very roughly smooth the times series using some sort of a moving average or moving median. By ‘very roughly’ I mean applying the minimum degree of smoothing possible (i.e. three periods), because the idea is just to eliminate short-run aberrations, and not to iron the short-run dynamics out of the series. The advantage of the median over the mean is that it completely eliminates fluke observations from the series, instead of just spreading the fluke over multiple observations.

Example 3: *Short-run noise in process C and short-run dynamics in process D generate identical entries in the observed transition matrix.* Applying a 3 period moving median to each of the series prior to the estimation of the transition probability matrix maintains the short-run dynamics in process within the series D, and filters the short-run noise in process C out of the series. The observed transition matrix does not change for D, whereas the off-diagonal transitions are wiped out from the one for C:

$$N_C^c = N_C^f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad n_C^c(\infty)' = n_C^f(\infty)' = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{23}$$

$$N_D^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, n_D^c(\infty)' = \begin{bmatrix} 0 \\ 0.9 \\ 0.1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow N_D^f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, n_D^f(\infty)' = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{24}$$

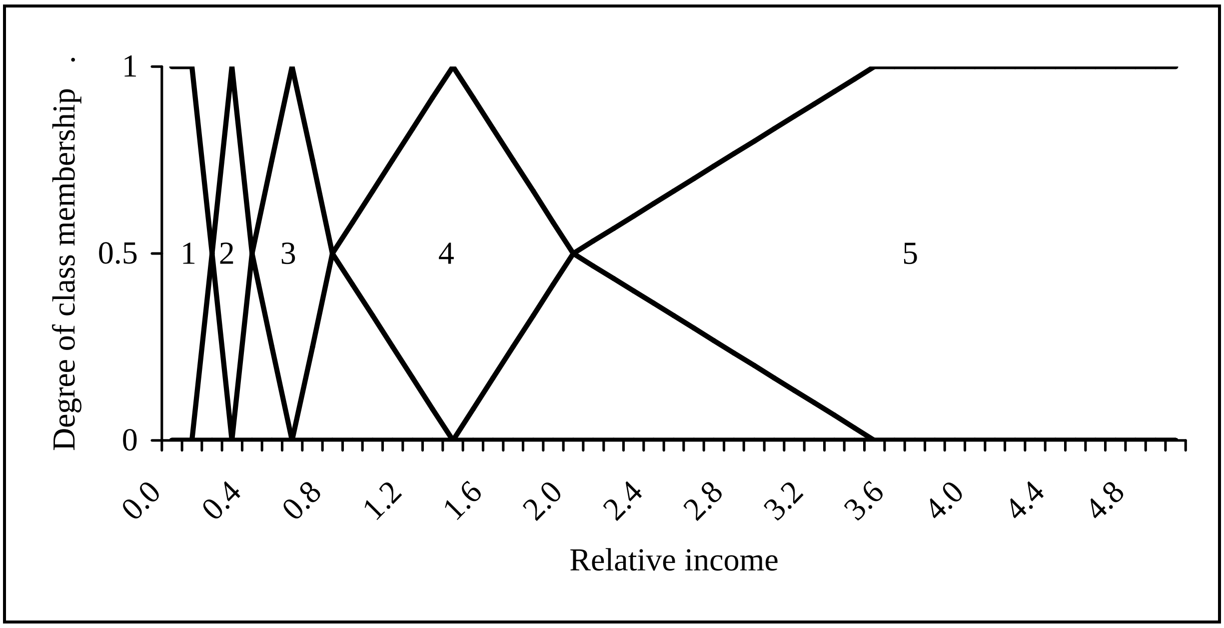
Notice that these results for the observed transition matrices can be equivalently expressed in terms of the ergodic distributions; corrections are carried out for C, whereas nothing changes for D.

6. Results

6.1 Fuzzification of the income class frontiers

Figure 12 depicts the maximum fuzzification of the income class frontiers that is carried out in this section.

Figure 12: Maximum fuzzification of the income class frontiers

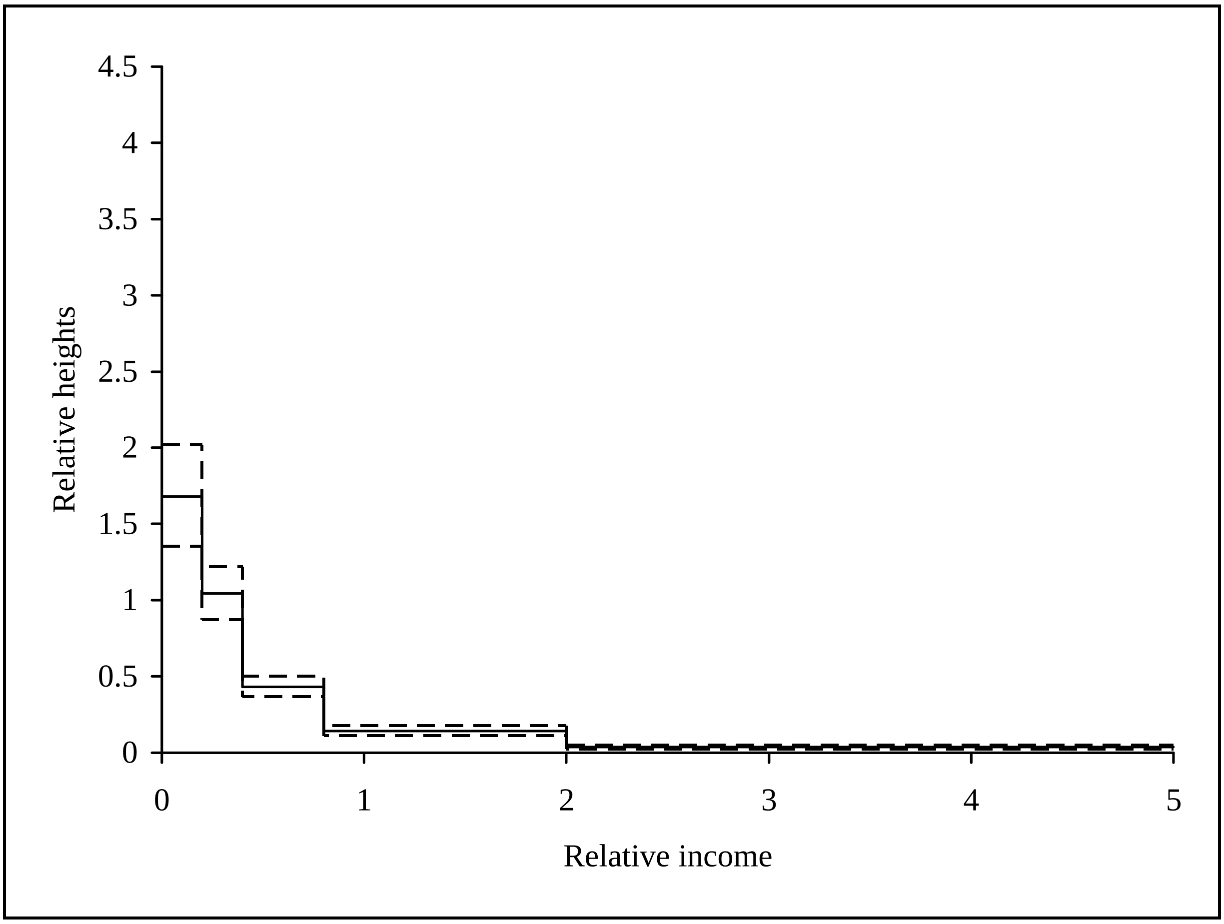


The estimation results are presented in Appendix A.2. The estimated fuzzy transition probability matrix is extremely similar to its crisp counterpart. Indeed, hypothesis testing indicates that the estimated fuzzy transition probability matrix falls within the confidence region calculated for its crisp counterpart. This similarity between the estimated transition matrices suggests that the incorrectly unaccounted intra-class mobility (c.f. Country B in Example 2) more or less compensates for the incorrectly accounted inter-class immobility (c.f. Country A in Example 1). The unscaled and rescaled fuzzy ergodic distributions not only fall within the confidence regions calculated for their crisp counterparts, they also display the same general shapes, twin peaks for the unscaled case and and poverty peak for the rescaled case.

What does fuzzification add to the standard Markovian analysis of the cross-country convergence issue? Efficiency. Whereas the crisp results are based upon only 156 total observed transitions, and contained within uselessly huge confidence regions (c.f. Appendix A.1), the fuzzy results are based upon 4077 observations of transitions[14], and contained within much tighter confidence regions (c.f. Appendix A.2). Indeed, the confidence region calculated for the rescaled fuzzy ergodic distribution is tight enough to render the poverty trap characterization of the ergodic distribution robust (c.f. Figure 13).

Figure 13: Histogram of long-run tendencies

(solid line: rescaled estimated fuzzy income group ergodic distribution; dashed lines: upper and lower bounds defining 95% confidence region)



[14] In the fuzzy analysis, the number of observed transitions (N) differs from the number of observations of transitions (N_{count}) because, contrary to the crisp analysis, observations can take rational values.

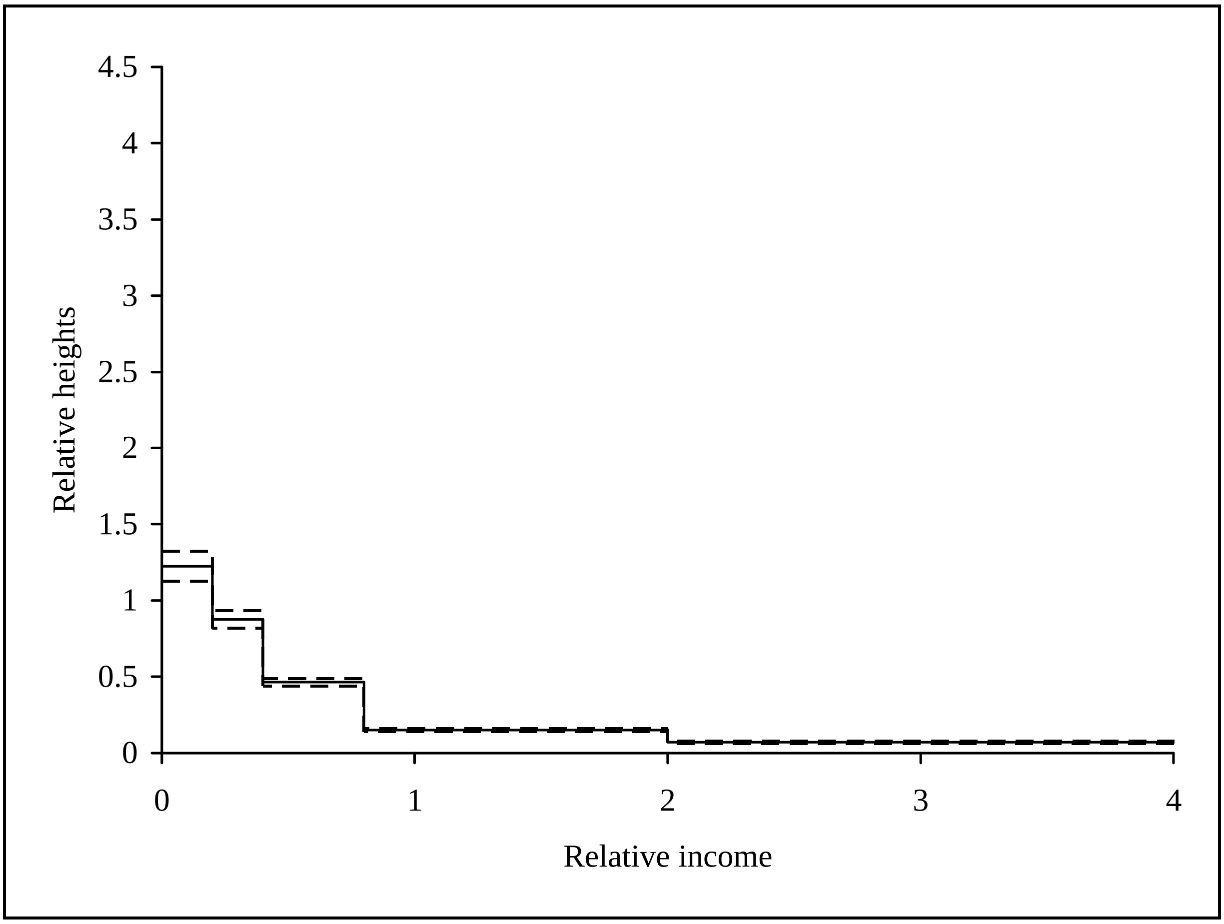
6.2 Fuzzification of the income observations

Fuzzification of the income observations is carried out using the partially subjective confidence intervals provided by Heston, Summers and Aten (2006) and Summers and Heston (1984, 1991, 1994, 2002). Grade “A” data is precise to more or less 5-10%, grade “B” data to 10-20%, grade “C” data to 20-30%, and grade “D” data to 30-40%. Values of 5%, 15%, 25% and 35% are used to carry out the calculations. The estimation results are presented in Appendix A.3. Although the fuzzy results remain similar to the crisp results, certain differences do appear. The estimated fuzzy transition probability matrix no longer falls within the confidence region calculated for its crisp counterpart, the fuzzy transition matrix presenting significantly more mobility than its crisp counterpart. The unscaled fuzzy ergodic distribution remains twin-peaked, but the peaks have lost in prominence. The rescaled fuzzy ergodic distribution continues to present its poverty peak.

Once again, fuzzification adds efficiency to the standard Markovian analysis of the cross-country convergence issue. Whereas the crisp results are based upon only 156 total observed transitions, and contained within uselessly huge confidence regions (c.f. Appendix A.1), the fuzzy results are based upon 3961 observations of transitions, and contained within much tighter confidence regions. As before, the confidence region calculated for the rescaled fuzzy ergodic distribution is tight enough to render the poverty trap characterization of the ergodic distribution robust (c.f. Figure 14).

Figure 14: Histogram of long-run tendencies

(solid line: rescaled estimated fuzzy income observation ergodic distribution; dashed lines: upper and lower bounds defining 95% confidence region)



As illustrated in Examples 1 and 2, fuzzification greatly changes (improves) the Markovian representation of a single country's path through the cross-country income distribution. Because of the Law of Large Numbers, however, fuzzification does not fundamentally alter the Markovian representation of aggregate world mobility. The contribution of fuzzy logic to the cross-country convergence application is to render the results of the simple Markov chain model statistically robust.

7. Concluding remarks

In the cross-country income convergence debate, the use of a discrete Markov chain model to approximate a continuous phenomenon introduces enough noise into the estimation process to undermine the robustness of results. The estimated transition probability matrix is biased and the resulting ergodic distribution fragile. Fuzzification of both the income class frontiers and the income observations provides two continuous adaptations of the discrete Markov chain model that improve the robustness of the results in a complementary fashion. Why fuzzification improves robustness can be understood in terms of statistical efficiency. The crisp Markov chain model can only extract binary information from the data, whereas the fuzzy Markov chain model can extract more refined information as well. Thus the fuzzy adaptation improves results by extracting information that is left behind by the crisp model from the noise and using it in a meaningful manner.

8. References

Adelman, I.G., 1958, A stochastic analysis of the size distribution of farms, Journal of the American Statistical Association 53, 893-904.

Aebi, R., K. Neusser, and P. Steiner, 2001, Where did the middle class go? An investigation into the dynamics of the US income distribution, Mimeo.

Aghevli, B.B., and F. Mehran, 1981, Optimal grouping of income distribution data, Journal of the American Statistical Association 76, 22-26.

Anderson, T.W., 1954, Probability models for analyzing time changes in attitudes In Lazarsfeld, P.F. (Ed.), Mathematical Thinking in the Social Sciences, Russell and Russell, New York.

Avrachenkov, K.E., and E. Sanchez, 2002, Fuzzy Markov chains and decision-making, Journal of Fuzzy Optimization and Decision-Making 1, 143-159.

Bianchi, M., 1997, Testing for convergence: evidence from non-parametric multimodality tests, Journal of Applied Econometrics 12, 393-409.

Blumen, I.M., M. Kogan, and P.J. McCarthy, 1955, The Industrial Mobility of Labor as a Probability Process, Cornell University Press, Ithaca.

Buchinsky, M., and J. Hunt, 1996, Wage mobility in the United States, NBER Working Paper 5455.

Bulli, S., 2001, Distribution dynamics and cross-country convergence: a new approach, Scottish Journal of Political Economy 48, 226-243.

Champernowne, D.G., 1953, A model of income distribution, The Economic Journal 63, 318-351.

Constant, A., and K.F. Zimmermann, 2003, The dynamics of repeat migration: a Markov chain analysis, CEPR Discussion Paper 4124.

Davies, J.B., and A.F. Shorrocks, 1989, Optimal grouping of income and wealth data, Journal of Econometrics 42, 97-108.

Dawson, J.W., J.P. DeJuan, J.J. Seater, and E.F. Stephenson, 2001, Economic information versus quality variation in cross-country data, Canadian Journal of Economics 34, 988-1009.

Fingleton, B., 1997, Specification and testing of Markov chain models: An application to convergence in the European Union, Oxford Bulletin of Economics and Statistics 59, 385-403.

Glass, D.V., and Hall, 1954, In Glass, D.V. (Ed.), Social Mobility in Britain, Routledge and Kegan Paul, London.

Heston, A., R. Summers, and B. Aten, 2006, Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices, University of Pennsylvania.

Johnson, P.A., 2000, A nonparametric analysis of income convergence across the US states, Economics Letters 69, 219-223.

Kemeny, J.G. and J.L. Snell, 1960, Finite Markov Chains, D. Van Nostrand Co. Inc., Princeton.

Koop, G., 2001, Modelling the evolution of distributions: An application to Major League Baseball, Edinburgh School of Economics DP 71.

Kremer, M., A. Onatski, and J. Stock, 2001, Searching for prosperity, NBER Working Paper 8250.

Landon-Lane, J.S., and J.A. Quinn, 2000, Growth and ergodicity: has the world converged?, Rutgers University, Mimeo.

Lavezzi, A.M., and D. Fiaschi, 2004, Nonlinear growth and the productivity slowdown, Computing in Economics and Finance 162.

Lee, T.C., G.G. Judge, and A. Zellner, 1977, Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North-Holland Publishing Company, Amsterdam.

Magrini, S., 1999, The evolution of income disparities among the regions of the European Union, Regional Science and Urban Economics 29, 257-281.

McFarland, D.D., 1970, Intragenerational social mobility as a Markov process: Including a time-stationary Markovian model that explains declines in observed mobility rates over time, The American Sociological Review 35, 463-476.

Pearlman, J.G., 2001, Twin peaks – A reassessment, Economics Department, London Guildhall University, Mimeo.

Prais, S.J., 1955, Measuring social mobility, Journal of the Royal Statistical Society Series A 118, 56-66.

Proudman, J., and S. Redding, 1998, Persistence and mobility in international trade In Proudman, J., and S. Redding (Eds.), Openness and Growth, Bank of England, London.

Proudman, J., S. Redding and M. Bianchi, 1998, Is international openness associated with faster economic growth? In Proudman, J., and S. Redding (Eds.), Openness and Growth, Bank of England, London.

Quah, D., 1993a, Empirical cross-section dynamics in economic growth, European Economic Review 37, 426-434.

Quah, D., 1993b, Galton's fallacy and tests of the convergence hypothesis, Scandinavian Journal of Economics 95, 427-443.

Quah, D., 1996a, Empirics for economic growth and convergence, European Economic Review 40, 1353-1375.

Quah, D., 1996b, Convergence empirics across economies with (some) capital mobility, Journal of Economic Growth 1, 95-124.

Quah, D., 1997, Empirics for growth and distribution: polarization, stratification convergence clubs, Journal of Economic Growth 2, 27-59.

Reichlin, L., 1999, Convergence as distribution dynamics: discussion In Baldwin, R., Cohen, D., Sapir, A., and A. Venables (Eds.), Regionalism and the Global Economy, Cambridge University Press, Cambridge.

Rogoff, R.N., 1953, Recent Trends in Occupational Mobility, The Free Press, Glencoe, Illinois.

Rummel, O., 2005, Distribution dynamics with measurement error, Monetary Instruments and Markets Division, Bank of England, Mimeo.

Shorrocks, A.F., 1976, Income mobility and the Markov assumption, The Economic Journal 86, 566-578.

Singer, B., and S. Spilerman, 1974, Social mobility models for heterogeneous populations, Sociological Methodology, Jossey-Bass, San Francisco.

Skaggs, R., and S. Ghosh, 1999, Assessing changes in soil erosion rates: A Markov chain analysis, Journal of Agricultural and Applied Economics 31, 611-622.

Summers, R., and A. Heston, 1984, Improved international comparisons of real product and its composition: 1950-1980, Review of Income and Wealth 33, 207-262.

Summers, R., and A. Heston, 1991, The Penn World Table (Mark 5): an expanded set of international comparisons, 1950-1988, The Quarterly Journal of Economics 106, 327-368.

Summers, R., and A. Heston, 1994, The Penn World Table (Mark 5.6): an expanded set of international comparisons, 1950-1992, National Bureau of Economic Research, Mimeo.

Summers, R., and A. Heston, 2002, The Penn World Table (Mark 6.0): an expanded set of international comparisons, 1950-1998, National Bureau of Economic Research, Mimeo.

Telser, L.G., 1962, Advertising and cigarettes, Journal of Political Economy 70, 471-499.

Temel, T., and P.J. Albersen, 2000, New facts for an old debate: Farm size, productivity and geography, Centre for World Food Studies, Mimeo.

Temel, T., and P.J. Albersen, 2001, Are the U.S. farm wages equalizing? Markov chain approach, Centre for World Food Studies, Mimeo.

Villaverde, J., and B. Sanchez-Robles, 2001, Convergence or twin peaks: the Spanish case, Departamento de Economia, Universidad de Cantrabria WP 2-01.

Appendix A.1: Results from the crisp application of the simple Markov chain model

$$N^c = \begin{bmatrix} 880 & 16 & 0 & 0 & 0 \\ 35 & 714 & 23 & 0 & 0 \\ 0 & 25 & 784 & 18 & 0 \\ 0 & 0 & 23 & 817 & 10 \\ 0 & 0 & 0 & 6 & 863 \end{bmatrix}$$

$$\hat{P}^c = \begin{bmatrix} 0.98 & 0.02 & 0 & 0 & 0 \\ 0.05 & 0.92 & 0.03 & 0 & 0 \\ 0 & 0.03 & 0.95 & 0.02 & 0 \\ 0 & 0 & 0.03 & 0.96 & 0.01 \\ 0 & 0 & 0 & 0.01 & 0.99 \end{bmatrix}$$

$$\hat{n}^c(\infty) = \begin{bmatrix} 0.38 & 0.15 & 0.15 & 0.12 & 0.20 \end{bmatrix}$$

$$\hat{n}^c_{rescaled}(\infty) = \begin{bmatrix} 1.90 & 0.75 & 0.37 & 0.10 & 0.07 \end{bmatrix}$$

$$\hat{P}^c \in \left\{ \begin{bmatrix} 0.974 & 0.009 & 0 & 0 & 0 \\ 0.031 & 0.906 & 0.018 & 0 & 0 \\ 0 & 0.019 & 0.933 & 0.012 & 0 \\ 0 & 0 & 0.016 & 0.948 & 0.005 \\ 0 & 0 & 0 & 0.001 & 0.988 \end{bmatrix}, \begin{bmatrix} 0.991 & 0.027 & 0 & 0 & 0 \\ 0.060 & 0.944 & 0.042 & 0 & 0 \\ 0 & 0.042 & 0.963 & 0.032 & 0 \\ 0 & 0 & 0.038 & 0.974 & 0.019 \\ 0 & 0 & 0 & 0.012 & 0.999 \end{bmatrix} \right\}$$

$$\hat{n}^c(\infty) \in \left\{ \begin{bmatrix} 0.017 & 0.009 & 0.016 & 0.011 & 0.006 \end{bmatrix}, \begin{bmatrix} 0.803 & 0.336 & 0.383 & 0.385 & 0.872 \end{bmatrix} \right\}$$

$$\hat{n}^c_{rescaled}(\infty) \in \left\{ \begin{bmatrix} 0.08 & 0.04 & 0.04 & 0.01 & 0.00 \end{bmatrix}, \begin{bmatrix} 4.01 & 1.68 & 0.96 & 0.32 & 0.29 \end{bmatrix} \right\}$$

Appendix A.2: Results from the fuzzy income class application of the simple Markov chain model

$$N^f = \begin{bmatrix} 856.03 & 18.80 & 0 & 0 & 0 \\ 31.22 & 851.76 & 20.87 & 0 & 0 \\ 0 & 25.45 & 866.98 & 17.79 & 0 \\ 0 & 0 & 19.10 & 940.13 & 14.19 \\ 0 & 0 & 0 & 12.44 & 539.24 \end{bmatrix}$$

$$\hat{P}^f = \begin{bmatrix} 0.98 & 0.02 & 0 & 0 & 0 \\ 0.03 & 0.94 & 0.02 & 0 & 0 \\ 0 & 0.03 & 0.95 & 0.02 & 0 \\ 0 & 0 & 0.02 & 0.97 & 0.01 \\ 0 & 0 & 0 & 0.02 & 0.98 \end{bmatrix}$$

$$\hat{n}^f = \begin{bmatrix} 0.34 & 0.21 & 0.17 & 0.17 & 0.11 \end{bmatrix}$$

$$\hat{n}^f_{rescaled} = \begin{bmatrix} 1.68 & 1.04 & 0.43 & 0.14 & 0.04 \end{bmatrix}$$

$$\hat{P}^f \in \left\{ \begin{bmatrix} 0.977 & 0.020 & 0 & 0 & 0 \\ 0.033 & 0.940 & 0.022 & 0 & 0 \\ 0 & 0.026 & 0.950 & 0.018 & 0 \\ 0 & 0 & 0.019 & 0.964 & 0.014 \\ 0 & 0 & 0 & 0.021 & 0.976 \end{bmatrix}, \begin{bmatrix} 0.980 & 0.023 & 0 & 0 & 0 \\ 0.036 & 0.945 & 0.025 & 0 & 0 \\ 0 & 0.030 & 0.955 & 0.021 & 0 \\ 0 & 0 & 0.021 & 0.967 & 0.016 \\ 0 & 0 & 0 & 0.024 & 0.979 \end{bmatrix} \right\}$$

$$\hat{n}^f(\infty) \in \{[0.271 \quad 0.175 \quad 0.146 \quad 0.133 \quad 0.079], [0.404 \quad 0.244 \quad 0.201 \quad 0.216 \quad 0.152]\}$$

$$\hat{n}^f_{rescaled}(\infty) \in \{[1.36 \quad 0.87 \quad 0.37 \quad 0.11 \quad 0.03], [2.02 \quad 1.22 \quad 0.50 \quad 0.18 \quad 0.05]\}$$

$$N^f_{count} = \begin{bmatrix} 1328 & 457 & 0 & 0 & 0 \\ 700 & 1903 & 380 & 0 & 0 \\ 0 & 480 & 1652 & 446 & 0 \\ 0 & 0 & 453 & 1994 & 637 \\ 0 & 0 & 0 & 524 & 1178 \end{bmatrix}$$

Appendix A.3: Results from the fuzzy income observation application of the simple Markov chain model

$$N^f = \begin{bmatrix} 820.76 & 100.79 & 0.39 & 0 & 0 \\ 115.03 & 533.53 & 101.85 & 0 & 0 \\ 0.06 & 106.41 & 628.06 & 90.82 & 0 \\ 0 & 0.01 & 95.09 & 702.23 & 47.79 \\ 0 & 0 & 0 & 42.02 & 829.15 \end{bmatrix}$$

$$\hat{P}^f = \begin{bmatrix} 0.89 & 0.11 & 0.00 & 0 & 0 \\ 0.15 & 0.71 & 0.14 & 0.00 & 0 \\ 0.00 & 0.13 & 0.76 & 0.11 & 0 \\ 0 & 0.00 & 0.11 & 0.83 & 0.06 \\ 0 & 0 & 0 & 0.05 & 0.95 \end{bmatrix}$$

$$\hat{n}^f = \begin{bmatrix} 0.25 & 0.18 & 0.19 & 0.18 & 0.21 \end{bmatrix}$$

$$\hat{n}^f_{rescaled} = \begin{bmatrix} 1.23 & 0.88 & 0.46 & 0.15 & 0.07 \end{bmatrix}$$

$$\hat{P}^f \in \left\{ \begin{bmatrix} 0.887 & 0.106 & 0.000 & 0 & 0 \\ 0.152 & 0.709 & 0.134 & 0.000 & 0 \\ 0.000 & 0.127 & 0.759 & 0.109 & 0 \\ 0 & 0.000 & 0.111 & 0.830 & 0.056 \\ 0 & 0 & 0 & 0.043 & 0.946 \end{bmatrix}, \begin{bmatrix} 0.894 & 0.113 & 0.001 & 0 & 0 \\ 0.155 & 0.713 & 0.137 & 0.000 & 0 \\ 0.000 & 0.130 & 0.763 & 0.111 & 0 \\ 0 & 0.000 & 0.113 & 0.832 & 0.057 \\ 0 & 0 & 0 & 0.054 & 0.957 \end{bmatrix} \right\}$$

$$\hat{n}^f(\infty) \in \left\{ \begin{bmatrix} 0.225 & 0.163 & 0.175 & 0.168 & 0.187 \end{bmatrix}, \begin{bmatrix} 0.265 & 0.187 & 0.195 & 0.195 & 0.243 \end{bmatrix} \right\}$$

$$\hat{n}^f_{rescaled}(\infty) \in \left\{ \begin{bmatrix} 1.13 & 0.82 & 0.44 & 0.14 & 0.06 \end{bmatrix}, \begin{bmatrix} 1.32 & 0.93 & 0.49 & 0.16 & 0.08 \end{bmatrix} \right\}$$

$$N^f_{count} = \begin{bmatrix} 1252 & 586 & 11 & 0 & 0 \\ 609 & 1265 & 595 & 5 & 0 \\ 13 & 608 & 1313 & 517 & 0 \\ 0 & 4 & 526 & 1195 & 249 \\ 0 & 0 & 0 & 238 & 987 \end{bmatrix}$$

Part II

The Dynamics of Intra-Household Time Allocation

Chapter 3

Nature versus Nurture:

An empirical analysis of the division of labor within American couples

Abstract: Why are women less active than men on the labor market? Is it choice or constraint? My results show that observed labor market participation does not reflect true preferences. I use PSID data for 2001 to estimate a simultaneous equations model in which time spent doing work around the house and on the labor market by 'husbands' and 'wives' are the four endogenous variables. In order to identify true preferences, I use four different proxies for the distribution of power between spouses. For wives, an increased share of the total education and wages for the couple is accompanied by an increased probability of working *longer* hours. For husbands, an increased proportion of women available on the remarriage market is accompanied by an increased probability of working *shorter* hours. These increases in probabilities are sufficient to alter the effective labor supply of both spouses.

Keywords: Marriage model; Time allocation; Housework; Labor supply; Power.

JEL classification: D13 ; J12 ; J16 ; J22.

1. Introduction

Fewer women than men participate in the US labor market. When women do participate, they work fewer hours on average than men (see Figures 1a and 1c). Why are women less active than men on the labor market? Is it nature or nurture, choice or constraint? My results show that observed labor market participation does not reflect true preferences.

Disentangling choice from constraint is difficult. To do this, I use intra-spousal bargaining power as a means of identifying the wife's freedom of choice. The more power a wife has relative to her husband, the more her actions will reflect her preferences. Four proxies for power are used: age, education, wage and local sex ratios. Using data from the Panel Study of Income Dynamics (PSID) on cohabiting couples for 2001, I carry out multinomial probit regressions of female labor market participation upon the four "power" variables. The results show that wives exploit education and wage power to work longer hours. The coefficients on the education and wage power variables are positive and significantly different from zero, meaning that an increased contribution by the wife to the total education and wages for the couple leads to an increased probability of her working longer hours. Conversely, wives without such power work shorter hours. In other words, the effective labor market supply of wives is constrained at levels that are below preferred ones.

For illustrative purposes, consider a "typical" woman with a college degree and the corresponding average wage. My results show that this same woman will most probably work full-time when she is more powerful than her husband, and part-time when she is less powerful. When this woman lives with a less powerful man (i.e. one with less education and a lower wage), she has a 26% chance of working part-time and a 44% chance of working full-time. When this woman lives with a more powerful man (i.e. one with more education and a higher wage), her chance of working part-time increases by 14% to reach 40%, and her chance of working full-time decreases by 11% to reach 33%. More generally, my results show that wives in more powerful situations are free to work more and that wives in less powerful situations are constrained to work less.

1.1 Female labor market preferences are important for policy

If the difference between the labor force participation of men and women is indeed due to constraint and not to choice, then what we are witnessing on labor markets is a violation of the basic human right to equity and a huge waste of economic efficiency. Such a ubiquitous market failure would require government intervention. Public inertia would continue to cost society in terms of lower GDP, reduced growth and diminished welfare.

From a microeconomic point of view, the optimal allocation of resources within a market economy relies upon competition and the barriers to entry to the labor market faced by women severely limit the scope and intensity of this competition. Moreover, in a market economy where two thirds of the consumers are women, a woman's personal know-how (in addition to her technical know-how) is a valuable resource. Indeed, it has been shown that when this resource is tapped, firms' financial results are higher.[15]

From a macroeconomic point of view, the ageing of the population is imposing an increasing strain on public resources (The Economist 2006), and the barriers to entry faced by women only serve to further burden these public resources. Increased female participation in the labor market would help alleviate this problem by expanding the tax base and reducing government transfers. Moreover, it would allow society to cash in on the return to the monumental investment made in the education of girls and women.

Castleman and Reed (2003) sum up the basic message in this subsection in a particularly poignant manner: "The significance of the debate between those who stress the social constraints on career-family choices and those who stress voluntarist decision-making and choice rests on the social implications for social policy at least as much as theoretical explanations of behaviour. If our social institutions and attitudes constrain women and men to limiting options, then we are bound to take action. If the outcomes are explained in terms of what people want to do anyway, then action is both futile and counterproductive" (p.2).

[15] For example, the U.S. Glass Ceiling Commission's 1995 Report (p.14) cites the results of a confidential study carried out by Covenant Investment Management: "Companies which rated in the bottom 100 on glass ceiling related measures earned an average of 7.9% return on investment, compared to an average return of 18.3% for the top 100." See also Rappoport etal. (2002) for a dynamic exploration of how improved work / private-life integration (not work-family balance) con foster better business outcomes and increase workplace performance and productivity.

1.2 Female labor market preferences are controversial.

Lisa Belkin's article in The New York Times Magazine on "The Opt-Out Revolution" and the nation-wide debate that it inspired illustrates the controversy surrounding the whole "choice versus constraint" question, and identifies key elements that must be considered when analyzing it. The main issue at stake was the identification of preferences. Belkin (2003) claimed that highly successful careerwomen with university degrees from top American universities are leaving the labor market in order to stay at home to care for their young children because "they don't want to [run the world]". Other journalists vigorously opposed her conclusions (Bauchner 2003, Douglas 2003, Pollitt 2003, Walsh 2003, Young 2004), highlighting instead the lack of alternatives: "Neither woman [discussed in Belkin's article] could get a part-time contract – it was 'all or nothing.' They didn't want to go back home, they wanted normal hours or, failing that, part-time jobs at decent salaries with real opportunities for advancement. If quitting was a 'choice,' it was a very constrained one" (Pollitt 2003).

This opposition between journalists mirrors a similar controversy between the general approaches taken in economics and sociology. Whereas both approaches are wary of the practical validity of people's *stated* preferences (in this case, those reported by the journalists), each approach deals with this problem differently. Empirical economists believe that true preferences are revealed by actions, and they use these *revealed* preferences to rationalize the status quo (i.e. if a woman does not work, it must mean that she does not want to work) (Kan 2005). In this sense, economists belong to the « nature » school. Sociologists question *revealed* preferences as much as *stated* preferences, viewing opinions and actions as constituting interdependent components of some sort of "coping mechanism" (i.e. some sort of ex-post justification of an action that an individual is "forced" to take ex-ante). They focus efforts on identifying the constraints shaping individuals' opinions and actions (Kan 2005). In this sense, sociologists belong to the « nurture » school. Duesenberry (1960) sums up the economics/sociology dichotomy in the following manner: "… the difference between economics and sociology is very simple. Economics is all about how people make choices. Sociology is all about why they don't have any choices to make" (p.233).

Even within the economics literature a similar controversy is found. Grossbard-Schechtman (2003) builds upon the basic belief that women prefer not participating in the labor market and finds evidence in favor of the opt-out theory. She constructs a theoretical model that "is most applicable to women who prefer not to work in the labor market" (p.16), and then demonstrates that current trends are consistent with her model. Goldin (2006) builds upon the basic belief that women prefer participating in the labor market and finds evidence against the opt-out theory. She postulates that "… most [women] perceive their work as a fundamental aspect of their satisfaction in life …" (p.12), and then demonstrates how the Current Population Survey and the College and Beyond datasets do not support any current trend towards an opt-out revolution.

1.3 General approach

These controversies demonstrate how difficult it is to pin down preferences. This is my first point. Another two fundamental issues relating to the female labor market participation decision, need to be discussed: the presence of labor market rigidities and the role of the husband within the household.

My second point is that labor market rigidities play a prominent role in the context of the opt-out debate. Most jobs just come in a couple of flavors (Martinez-Granado 2005): part-time and full-time. As we move up the career ladder, more and more jobs come in the single take-it-or-leave-it, whatever-it-takes-to-get-done, full-time flavor. Going back to Figures 1a and 1c, we see that 42% of wives work either 0 or 40 hours a week and that 44% of husbands work either 40, 45 or 50 hours a week.

My third point is that the role of the husband in the process of household time allocation is strikingly absent from the whole opt-out debate. Rhode notes this absence and comments: "If women are not choosing to run the world it is because men are not choosing to run the washer dryer" (Burk, Gillette & Rhode 2005). Implicit in this remark is the application of a very particular model of household time allocation to interpret the opt-out evidence. First, husbands choose their contributions to housework and to market work. Second, wives do whatever remains to be done around the house (the husband's housework choice thus acts as a constraint upon the wife's housework choice). Third, wives do whatever they have time left to do on the labor market (the housework choices act as a constraint upon the wife's market work choice).

Combining these three issues in a potentially provocative manner for illustrative purposes, I construct the following hypothetical scenario. The husband decides not to run the washer-dryer (i.e. he does not make a substantial contribution to housework) and the wife is thus allocated the task of running the washer dryer (i.e. the bulk of the housekeeping activities). This housework occupies her for 20 hours a week. Given this burden, she does not want to work more than 40 hours a week on the labor market. But given her qualifications, the only jobs available to her are 60 hours a week. And so she is faced with the all-or-nothing choice of doing a "double-shift" or becoming a stay-at-home mom. Actually there is also a third option; she could leave her spouse (c.f. cooperative bargaining models, i.e. Manser & Brown 1981 and McElroy & Horney 1981). Separation or divorce would provide the researcher with a clear signal of preferences. Unfortunately (at least for the researcher), dissatisfaction with the intra-spousal distribution of housework does not systematically result in marital dissolution. Two other outcomes are just as compatible with the same underlying preferences. The wife could remain married, but unhappily so (c.f. non-cooperative bargaining models, i.e. Lundberg & Pollack 1995 / 1996). Or, the wife could construct some sort of coping mechanism allowing her to remain married, and happily so (Hochschild 1989). We are back to the problem of preference identification.

In this paper, I model the process of household time allocation in the most general manner possible. Each of the four time variables (the time spent doing house or market work by the wife or the husband) is allowed to depend upon the other three time variables. For example, the time spent doing housework by the wife is regressed upon the time spent doing housework by the husband, the time spent doing market work by the wife and the time spent doing market work by the husband. Labor market rigidities are incorporated into the model by defining the time spent doing market work by each of the spouses as discrete variables.

In the next section, I go into more detail about the identification strategy adopted in this paper, and in the following section, I present a brief review of the literature. The empirical model is presented in Section 4, the dataset and variables in Section 5, and the empirical findings and discussion in Section 6. The last section concludes.

2. Identification

2.1 Power

Choices result from the maximization of utility, subject to constraints. When there are many constraints, the choice reveals primarily information about the constraints. When there are fewer constraints, the choice reveals more information about the preferences. This basic idea in economics is formulated in terms of power in sociology. According to Max Weber, "… power is … every opportunity/possibility existing within a social relationship, which permits one to carry out one's own will, even against resistance, and regardless of the basis on which this opportunity rests." I exploit this semi-tautological relationship between preferences and power to turn a tricky preference identification problem into a simpler power identification problem.

So, what is power? I adopt the view implicit in Weber's definition that power is a very broad notion that can take many different forms. These forms can be material or psychological, and power is different for different people. Whereas in some couples money might tip the balance of power in favor of the spouse earning the money, in other couples fear of loneliness might tip the balance of power away from the spouse fearing loneliness.

In this paper, I use four proxies for power: age, education, wage and sex ratios. The first three proxies for power, age, education and wage, all contribute to representing the material aspect of the intra-spousal distribution of power. Age is a tricky variable because on the level of wisdom it generally favors the individual, but on the level of beauty the effect tends to be the opposite. I choose to think of age in terms of the labor market, in terms of providing the older spouse with a first-mover advantage over the younger spouse. The idea is that when the younger spouse moves into the labor market, his/her choice is made given whatever his/her spouse has already chosen to do. Whether or not this constraint is binding is one of the focuses of this paper. I include education and wage to represent potential and effective access to resources. I am not interested in the absolute levels of these variables, but how having more or less of it than your spouse constrains or enables your choices. The last proxy for power, sex ratios, contributes to capturing the psychological aspect of the intra-spousal distribution of power. This variable measures the relative abundance of potential new partners in the event of a break-up; it measures relative power on the remarriage market.

2.2 Specialization and exchange

If you are an economist, you might be thinking that the power story in general, and the wage power story in particular, is just a specialization and exchange story in disguise. In this section, I would like to clarify why you should not expect this to be the case ex-ante (i.e. before the econometric analysis), and why this proves not to be the case ex-post (i.e. after the econometric analysis). To be absolutely clear right from the start, the results for the wives are compatible with the specialization and trade story, but the results for the husbands are not.

So, how can I distinguish the power story from the specialization and exchange story? According to the specialization and exchange story, the spouse with the higher wage specializes in market work, and the other spouse specializes in housework. Note that this story is genderless. In other words, if this theory is true, then there are only two possible scenarios. In the first scenario, the husband earns a higher wage than the wife, so he specializes in market work, leaving her to specialize in housework. In the second scenario, the wife earns a higher wage than the husband, so she specializes in market work leaving him to specialize in housework. If empirical observations reveal that intra-spousal time allocation cannot be described by one of these two scenarios (i.e. if intra-spousal time allocation proves to be gendered), then the specialization and exchange story cannot be used to describe intra-spousal time allocation. In this case, the power story provides an alternative interpretation of the evidence.

In this paper, both descriptive and inferential statistics reveal much behaviour that confirms the first scenario and some behaviour that contradicts the second scenario. That husbands who earn the higher wage specialize in market work and that wives who earn the lower wage specialize in housework comes as no surprise (see Figure 1).

Figure 1: Intraspousal time allocation when the husband earns a higher wage than the wife
(Note: The outcome predicted by specialization and trade theory is indicated in bold.)

weekly market work hours \ weekly housework hours	wife<husband	wife=husband	wife>husband
wife<husband	3%	***8%***	***68%***
wife=husband	1%	***2%***	***6%***
wife>husband	2%	2%	9%

In 73% of the couples, the husband earns a higher wage than his wife. In this context, the specialization and exchange story points to the first scenario. The outcomes corresponding to this scenario are indicated in bold. Note that 84% of these couples act in accordance with the first scenario. So far, so good. But the specialization and exchange story breaks down when it is the wives who earn higher wages than their husbands (see Figure 2).

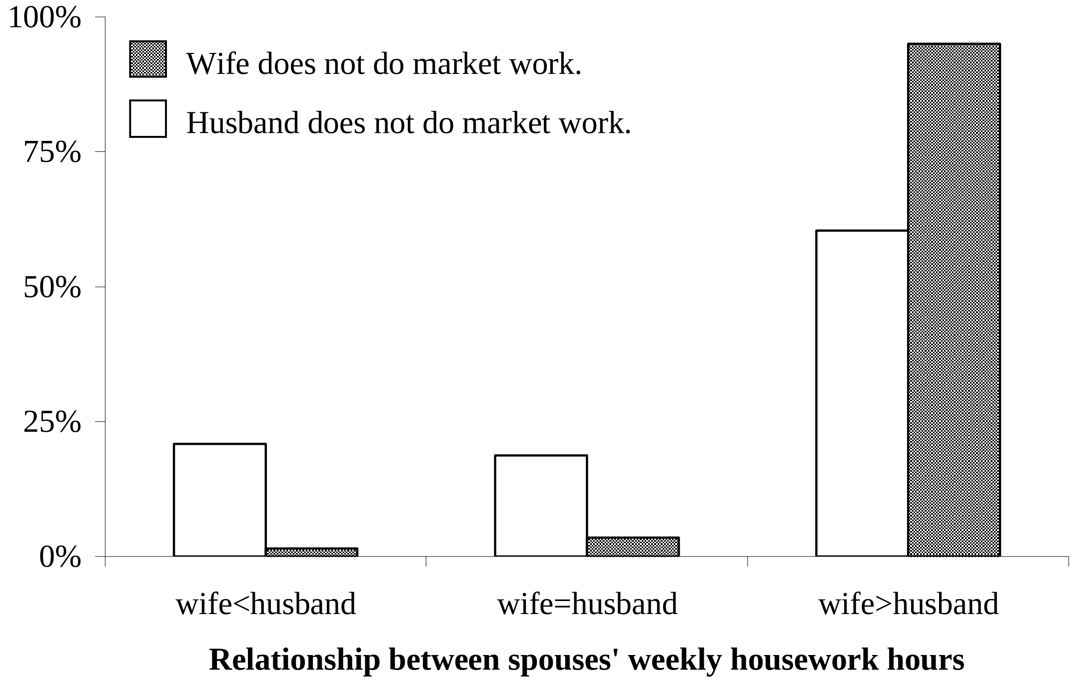
Figure 2: Intraspousal time allocation when the wife earns a higher wage than the husband
(Note: The outcome predicted by specialization and trade theory is indicated in bold.)

weekly market work hours \ weekly housework hours	wife<husband	wife=husband	wife>husband
wife<husband	3%	9%	47%
wife=husband	**2%**	**3%**	10%
wife>husband	**5%**	**5%**	17%

In 27% of the couples, the wife earns a higher wage than her husband. In this context, the specialization and exchange story points to the second scenario. The outcomes corresponding to this scenario are indicated in bold. Note that only 15% of these couples act in accordance with the second scenario. Indeed, this second intra-spousal time allocation matrix looks rather similar to the first one, meaning that wages do not seem to matter for the distribution of tasks.

This gendered intra-spousal allocation of time is perhaps even more evident for couples in which one of the spouses does not do any market work. In this case, specialization and exchange theory predicts that the stay-at-home spouse does more housework than the breadwinner spouse. Once again, this theory is verified when it is the wives who stay at home, but not when it is the husbands who stay at home (see Figure 3).

Figure 3: Histogram of intra-spousal housework time allocation when one spouse does not do market work



95% of the stay-at-home wives compensate for their unemployment with a relatively higher contribution to household production. The equivalent number for the stay-at-home husbands is only 21%. 1% of the husbands with stay-at-home wives contribute more to household production than their wives. The equivalent number for the wives of stay-at-home husbands is 60%.

In sum, it seems that most wives do more housework than their husbands, whether or not wives earn higher wages than husbands.[16] Of course, these are only descriptive statistics, so I still need to explicitly hold constant a whole host of other factors in order to conclude anything from this information. When I do hold these other variables constant, as is done in the empirical analysis, these results are maintained. The point that I would like to make in this section is just that this paper presents evidence that raise doubts about the validity of the specialization and exchange story, and explores an alternative explanation, a story about power.

[16] This asymmetry is somewhat of a stylized fact in Sociology, and the explanation is correspondingly asymmetrical. Exchange theory, as sociologists call specialization and exchange theory, is used to explain time allocation behavior for couples in which the husband earns more than the wife does, and gender theory is used to explain time allocation behavior for the rest of the couples. I argue that for specialization and exchange theory to hold, intra-spousal time allocation behavior cannot be gendered.

3. Literature review

3.1 Socio-economic approaches to preferences

Most of the socio-economic literature on the household time allocation problem does not focus on preferences themselves, but rather on the household decision-making process (for reviews, see Mattila-Wiro 1999, Anxo & Carlin 2004 and Pollack 2005). Two notable exceptions are economist Becker and sociologist Hakim. Becker places preferences beyond (economic) theory (Stigler & Becker 1977, Godwin 1991), whereas Hakim places preferences at the center of (sociological) theory (Hakim 2000). Whereas Becker and Hakim position themselves in diametric opposition, mainstream economics and sociology meet somewhere in the middle. With this paper, I position myself upon this middle ground.

In Sociology, Hakim takes an extreme and controversial stance by proposing that in modern societies[17], "women have genuine choices and female heterogeneity is revealed to its full extent" (Hakim 2000: 22). She provides highly contested evidence that stated preferences correspond to revealed preferences (Bruegel 1996, Ginn etal. 1996, Procter & Padfield 1999, McRae 2003) and she concludes that these preferences correspond to true preferences. Preference theory thus mitigates the role of constraints, thereby placing it in stark opposition to the rest of the sociological literature, which focuses on constraints and not on preferences (Kan 2005).

In Economics, Becker takes an equally extreme and controversial stance by proposing that "one may usefully treat tastes as stable over time and similar among people" (Stigler & Becker 1977: 76). In other words, Becker says that tastes do not matter. However, subsequent research has shown that tastes do matter, especially for welfare.

Becker's proposition constitutes the basis of his unitary household time allocation model in which the distribution of the spouses' time between household and market

[17] For Hakim (2000), a modern society is one in which the following five conditions are fulfilled: occurrence of the contraceptive revolution, occurrence of the equal opportunities revolution, expansion of white-collar occupations, creation of jobs for secondary earners, and increasing importance of values in lifestyle choices.

production activities results from the maximization of one single utility function. Much of the economic literature in the field has been spent refuting the unitary model in favor of collective models (cooperative and non-cooperative bargaining models) that take into account both spouses' utility functions. So, tastes do matter.

Moreover, Pollack raises an additional issue of particular interest here in his response to Becker: "For positive analysis, whether we attribute differences in behavior to unobserved differences in household technology rather than to unobserved differences in taste is mere semantics. For welfare analysis, however, whether we attribute differences in behavior to differences in technology rather than to differences in tastes can alter conclusions about whether a policy change increases or decreases welfare" (Pollack 2003:116). So, tastes do matter, especially for welfare.

Following Pollack's reasoning one step further, we see that economic and sociological thought meet in the middle ground between Becker and Hakim: "Variable tastes undermine the normative significance of the fundamental theorem of welfare economics which asserts … that in competitive equilibrium everyone gets what he wants … However, if tastes are sufficiently malleable, then this may be no more than a corollary of the more general proposition that people come to want what they get" (Pollack 1978:374, McCrate 1988). This description corresponds to what sociologists call "coping mechanisms", those ex-post justifications that individuals use to rationalize actions that they were "forced" to take ex-ante (Hochschild 1989). I position my paper in this middle ground shared by economics and sociology, where stated and revealed preferences meet and mingle, necessitating an identification strategy to disentangle the two.

3.2 Evidence on preferences

The economic literature providing empirical results on women's labor market preferences is sparse. A couple of papers that do present such evidence for the United States identify preferences using divorce and marital property legislation. Legislation on divorce and marital property affects what each spouse can expect to walk away with in the case of divorce thus affecting each spouse's external threat point and the balance of power between the spouses. Results are mixed. Chiappori, Fortin and Lacroix (2002) use cross-sectional data within a collective framework to show that in states where the divorce and marital property laws are favorable to women, female labor supply is lower that in other states. Gray (1998) uses a dynamic approach within a bargaining framework to show that changes in the divorce and marital property laws that were favorable to women generated an increase in female labor supply.

4. Empirical model

I do not want to impose any particular model of the time allocation process upon the data, so I choose to take Ghysel's (2003) approach which is general enough to nest the two main family labor supply models (the unitary and collective models). Each of the four time variables (the time spent doing house or market work by the wife or the husband) could potentially depend upon the other three, so these four variables are modeled as conditional labor supply functions and estimated within a simultaneous equations framework.

My starting point is the household's utility function V, which combines each spouse's individual utility U_i in a way that depends upon the distribution of power between them a :

$$V = V(U_w, U_h, a)$$

where w indicates wife and h indicates husband. Each individual spouse's utility function depends upon own consumption x_i, spouse's consumption x_j, household production H, own leisure l_i, spouse's leisure l_j, and personal characteristics g_i:

$$U_i = U_i(x_i, x_j, H, l_i, l_j, g_i)$$

Spouses maximize their personal functions subject to the (i) budget constraint, (ii) household technology constraint, and (iii) time constraint:

(i) $$px_i = f(w_i m_i, w_j m_j, y_i, y_j, y, a, g_i)$$

where p is the general price level, w_i is own wage, w_j is spouse's wage, m_i is own market work time, m_j is spouse's market work time, y_i is own non-labor income, y_j is spouse's non-labor income, and y is household non-labor income;

(ii) $$H = H(h_i, h_j, b)$$

where h_i is own housework time, h_j is spouse's housework time, and b is determinants of household production dependant upon couple characteristics;

(iii) $$T = m_i + h_i + l_i$$

where T is total time available for different activities (i.e. 16 or 24 hours, depending upon whether you count sleep as leisure or not). Incorporating these constraints into the individual utility functions, I can rewrite the household utility function in the following more informative manner:

$$V = V(m_w, m_h, h_w, h_h, w_w, w_h, y_w, y_h, y, a, b, g_w, g_h)$$

Using this definition of the household utility function within Pollack's (1969, 1971) framework of conditional demand functions, I derive the demand functions for house and market work time for each of the spouses (see Lundberg 1988 and Ghysels 2003 for similar applications of the conditional demand approach within the labor supply context). The idea here is to allow each spouse's time allocation decision in one sphere (house or market) to potentially depend upon their own time allocation in the other sphere and the other spouse's time allocations in both spheres. More concretely, I maximize the household utility function relative to each of the four time variables separately, conditional upon the values taken by the remaining three time variables. For spouse i, this gives me:

$$\max V_i = V_i(m_i, \overline{m}_j, \overline{h}_i, \overline{h}_j)$$
$$\max V_i = V_i(\overline{m}_i, \overline{m}_j, h_i, \overline{h}_j)$$

where $^{-}$ indicates the pre-allocated values. This maximization procedure generates a set of four demand functions. For spouse i, this gives me:

$$m_i^* = m_i(w_i, w_j, y_i, y_j, y, \overline{m}_j, \overline{h}_j, \overline{h}_i, a, b, g_i, g_j)$$
$$h_i^* = m_i(w_i, w_j, y_i, y_j, y, \overline{m}_j, \overline{h}_j, \overline{m}_i, a, b, g_i, g_j)$$

It is this system of four simultaneous equations that is the object of estimation. I take an equation-by-equation approach, estimating the continuous housework time equations using OLS and the discrete market work time equations using ordered probit. Such an approach necessitates a two-stage estimation procedure.

In the first stage, instrumentation is carried out for husbands' and wives' wages and the four time variables. These six variables are all endogenous variables so they need to be instrumented in order to include them in the regressions on the right-hand side. Also, wages are not observed for those people who do not work, so wages need to be assigned to these people.

In the second stage, the set of four equations is estimated. Econometric identification of the system of simultaneous equations is achieved via a combination of functional form and exclusion restrictions. Using OLS for the housework time equations imposes a linear functional form upon the estimators of the coefficients in those equations, and using the probit estimation procedure for the market work time equations imposes a nonlinear functional form upon the estimators of the coefficients in those equations; this ensures identification of the housework time equations. Including a squared term for the wives' housework time in the wives' market work time equation further contributes to the identification of the wives' housework time equation. As for the exclusion restrictions, I impose two. First, I assume that not all of the couple characteristics are included in the market work equations. Indeed, the number of rooms and the number of cars corresponding to a couple might have a direct impact upon housework hours, but not upon labor market hours. This exclusion restriction ensures identification of the market work time equations. Second, I assume that whereas all of the personal characteristics contained in g_i are included in each of the equations, only some of these characteristics for the spouse are included in the different equations. This exclusion restriction enhances overall identification of the system as a whole.

Before moving on to the more empirical part of this paper, I need to explicit how theory would manifest itself within this empirical framework. For the specialization and exchange story, the answer is easy. In the housework equations, the coefficients on the less-wage-power dummy need to be positive, and the coefficients on the more-wage-power dummy negative. In the market work equations, the coefficients on the less-wage-power dummy need to be negative, and the coefficients on the more-wage-power dummy positive. For the power story, I have no answer. The whole point of such a flexible econometric framework is to let the data reveal to us the true preferences of husbands and wives. The risk is that the results are compatible with the specialization and exchange story, situation in which it will not be possible to distinguish between the alternative explanations.

5. Dataset and variables

The data comes from the Panel Study of Income Dynamics (PSID) for the interview year 2001. This dataset provides information on a representative sample of the American population. My sample is composed of cohabiting couples residing within the US, in which both 'spouses' are between the ages of 25 and 59, endpoints included. To simplify communication, I refer to the partners of these co-habiting couples as 'husband' and wife', even though they might not be married. I exclude younger and older couples to minimize the impact of any specific behaviour related to studies and to retirement. The data is obtained through an interview with one of the two spouses, and I limit my sample to couples for which the data on the variables required is complete (i.e. couples with 'Don't Know' or 'Not Applicable' responses are eliminated from the sample). This gives me 1788 observations.

5.1 Housework and market work: m_h, m_w, h_h and h_w

The four main variables are the weekly hours spent by each of the spouses doing work around the house or on the labor market (***housework*** and ***marketwork***, respectively). The number of hours spent on housework every week comes from the answer to the following question: "About how much time do you spend on housework in an average week? (I mean time spent cooking, cleaning and doing other work around the house.)." The number of hours spent on the labor market every week is calculated by dividing the "Total hours of work in

2000" by the number of "Work weeks in 2000". The annual hours of work is itself a synthetic variable calculated by PSID using the following equation: "Work weeks on main job x Work hours per week + Overtime work hours + Extra work hours."

The accuracy of the data on these four main variables is certainly questionable. The definition supplied of housework is vague, so tasks that one respondent includes, another might not. Furthermore, the answer to the housework question relies upon accurate and objective recall as well as truthful reporting. Accurate recall is already difficult in and of itself given the irregular timing of at least some housework, and when you take into account the generally unpleasant nature of housework, time spent doing housework has a funny way of seeming interminably long. Housework has the additional characteristic of being a socially desirable activity, so since people generally do not want to be considered a slob or a free-rider, there is an incentive for respondents to exaggerate their contributions. The data on labor market hours should not suffer from these same potential inaccuracies. Hours worked on the labor market tend to be regular and are often explicitly defined in a contract. The problem here lies in the fact that this variable is the result of a calculation based upon five different variables, and that there are therefore five potential sources of error.

Despite these inaccuracies, I choose to work with PSID data for two reasons. First, to my knowledge there is not a better alternative. American Time Use Survey data would be preferable because respondents note what they are doing in a diary when they are doing it. However, such data is not collected for couples, but only for one spouse per couple. Since I model the times spent by each spouse working around the house or on the labor market as the outcome of a household decision making process, I need data for both spouses of each couple.

Second, studies on the inaccuracies present in PSID data provide information on the sign and magnitude of these biases, and confirm the validity of using survey data to study household time allocation. Juster, Ono and Stafford (2003) compare these four PSID variables to their diary-based counterparts. Their data shows that for market work, men tend to over-report hours by more or less 10%, whereas women tend to under-report hours by a similar margin.[18] For housework, both men and women inflate reports of their hours, women by around 40%, and men by 20-100% depending on the year. These results on housework are

[18] These results are based upon calculations that I carried out using the data presented in Table B1 of Juster, Ono and Stafford (2003), and they actually contradict what is presented in the paper.

roughly in line with the rest of the literature on the quality of survey data relative to time-use data in the US (Marini & Shelton 1993, Robinson, 1985, Bianchi, Milkie, Sayer & Robinson 2000). The general conclusion is that despite differences between survey and time-use data, survey data provides an ordinal scaling of individuals' time spent on housework that is useful for multivariate analyses of household time allocation (Kan 2006).

In this context of noisy data, it is particularly important to clean the data of outliers. Given the multivariate nature of the household time allocation problem, simple univariate or bivariate trimming of the dataset is not effective.[19] Here cleaning the data means detecting outliers in a multivariate point cloud when there are most probably several outliers. Intuitively, we need to identify the center of the cloud of observations, measure the distance of each observation from this center taking into account the shape of the cloud, and eliminate those observations that are 'too far' from the center. Rousseeuw and Van Zomeren (1990) propose the following robust version of the Mahalanobis distance:

$$RMD_i = \sqrt{(x_i - T(X)C(X)^{-1}(x_i - T(X))^t}$$

where x_i is the i^{th} observation of the dataset X, $T(X)$ is the center of the minimum volume ellipsoid estimator covering half of the observations and $C(X)$ is the sample covariance matrix determined by the same ellipsoid.[20] The cutoff distance is supplied by the square root of the appropriate value of the chi-square statistic.

I use the *lts* command in S-Plus to calculate the robust Mahalanobis distances first for a reduced version of the dataset containing only the four dependent variables and then for the complete version of the dataset containing all of the variables. Both plots clearly reveal the presence of a subpopulation (a cluster of observations located beyond the cutoff distance), composed of those couples in which the husband is unemployed (see Figures 2a and 2b). Even though these couples visibly behave differently from other couples, they are not outliers, so I do not want to eliminate them from the sample. Thus, I calculate the robust Mahalanobis

[19] Consider the following values for the four endogenous time variables: wife's housework hours=30, husband's housework hours=15, wife's market work hours=60, husband's market work hours=70. Neither of these values taken separately would qualify as an outlier, but all of these values taken together define a couple with atypical behavior, one that should be qualified as an outlier.

[20] The classical Mahalanobis distance defines *T(X)* as the arithmetic mean of *X* and *C(X)* as the usual sample covariance matrix. When these definitions are used the distance measure suffers from the "masking effect" by which multiple outliers do not necessarily present a large MD_i.

distances for the two populations separately, for both the reduced and the complete versions of the dataset in both cases, and I eliminate the outliers thus identified (see the observations marked in black in Figures 3a to 3d). This leaves me with 1516 observations.

The last thing that I need to do with these variables is to discretize the continuous market work time variables. The idea here is to account for labor market rigidities. When a spouse is deciding upon his/her market work contribution, (s)he faces limited options in terms of market work hours (Martinez-Granado 2005). So, when the market work time variables are on the left-hand side, they will be defined in categorical terms: no-time, part-time, full-time, over-time. When a spouse is deciding upon his/her housework contribution, this time allocation will be made based upon the exact number of hours dedicated to market work. Using the categorical definition of market work time in the housework equations would be wasting useful information. So, when the market work time variables are on the right-hand side, they will be defined in continuous terms.

Since the idea is to account for labor market reality, I use the empirical distribution of the population across the continuous market work time variables to guide the discretization of these variables. These distributions are presented in Figures 1a and 1c. Both distributions reveal a highly rigid labor market, but wives seem to benefit from more options in terms of market work hours than husbands do. Husbands seem to face a choice between full-time and over-time jobs, and wives between no-time, part-time, full-time and over-time jobs, so I define the following market work dummies:

$$job1 = \begin{cases} 0 & \text{if } \mathit{marketwork} = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$job2 = \begin{cases} 0 & \text{if } (\mathit{marketwork} \geq 0)\ \&\ (\mathit{marketwork} < 40) \\ 1 & \text{otherwise} \end{cases}$$

$$job3 = \begin{cases} 0 & \text{if } (\mathit{marketwork} \geq 40)\ \&\ (\mathit{marketwork} < 45) \\ 1 & \text{otherwise} \end{cases}$$

$$job4 = \begin{cases} 0 & \text{if } \mathit{marketwork} \geq 45 \\ 1 & \text{otherwise} \end{cases}$$

Only *job4* is relevant to the husbands, whereas the whole set of dummies is relevant to the wives.

5.2 Power: *a*

I use multiple measures of the intra-spousal distribution of power in order to capture as many aspects as possible of the very wide view of power that I have taken. The intra-spousal distribution of power is defined as an index in the following manner:

$$power\ index_{wife} = \frac{power_{wife}}{power_{wife} + power_{husband}}$$

This index varies from 0 to 1, 0 representing no power for the wife and total power for the husband, and 1 representing total power for the wife and no power for the husband. 0.5 represents equality between the two spouses.

I use four measures of power: age, education, wage and sex ratios. Age is measured in years. Education is the actual grade of school completed. 17 is the maximum value of the education variable and indicates that the individual has undertaken at least some postgraduate work. Wage is the average hourly remuneration in dollars (i.e. taking into account different wages for different jobs and extra bonuses paid in addition to the usual hourly rate of pay). The sex ratio for the wife is the ratio between the number of men and women residing within the same state and having the same age plus or minus 5 years. Note that I include married people in my definition of the remarriage market. Indeed, if I am talking about remarriage, then marriage is not binding, and if marriage is not binding then the remarriage market must include married people. The sex ratio for the husband is defined in the equivalent manner. The data used to calculate the sex ratios comes from Census data for 2001.

In order to allow for a maximum of flexibility in the modelling of the relationship between the dependent variables and these four power indexes (i.e. to allow for the possibility of different slopes), I transform each power index into a dummy variable with three possible values: husband is (relatively) more powerful, spouses share power (relatively) equally, and wife is (relatively) more powerful. Theoretically, all four indexes should be discretized in the same manner, using 0.5 as representative of equality. However, three related empirical considerations lead me to adapt this theoretical approach. First, the value of 0.5 representing exact equality is only practically feasible for the discrete variables, the age and education indexes. An interval including 0.5 would be more appropriate for the continuous variables, the wage and sex ratio indexes. Second, from the point of view of inference, it would be preferable to have a minimum number of observations for each of the three values. Third, the intervals need to be wide enough in order to identify groups with distinguishably different behaviours. Indeed, will a spouse with an observed power value of 0.51 really behave differently from one with a value of 0.50 and one with a value of 0.49?

These three considerations lead me to discretize the four indexes in the following way. The education index is discretized as theoretically desired, using 0.5 to represent equality (see Figure 4b):

$$edupwrlo = \begin{cases} 1 & \text{if } edupwr < 0.5 \\ 0 & \text{otherwise} \end{cases}$$
$$edupwrmed = \begin{cases} 1 & \text{if } edupwr = 0.5 \\ 0 & \text{otherwise} \end{cases}$$
$$edupwrhi = \begin{cases} 1 & \text{if } edupwr > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

The age and sex ratio indexes are discretized by defining an interval that is roughly centered upon 0.5 to represent equality (0.48-0.51 for the age index and 0.46-0.51 for the sex ratio index, see Figures 4a and 4c):

$$agepwrlo = \begin{cases} 1 & \text{if } agepwr < 0.48 \\ 0 & \text{otherwise} \end{cases}$$
$$agepwrmed = \begin{cases} 1 & \text{if } (agepwr \geq 0.48)\ \&\ (agepwr < 0.51) \\ 0 & \text{otherwise} \end{cases}$$
$$agepwrhi = \begin{cases} 1 & \text{if } agepwr \geq 0.51 \\ 0 & \text{otherwise} \end{cases}$$

and

$$sexpwrlo = \begin{cases} 1 & \text{if } sexpwr < 0.46 \\ 0 & \text{otherwise} \end{cases}$$
$$sexpwrmed = \begin{cases} 1 & \text{if } (sexpwr \geq 0.46)\ \&\ (sexpwr < 0.51) \\ 0 & \text{otherwise} \end{cases}$$
$$sexpwrhi = \begin{cases} 1 & \text{if } sexpwr \geq 0.51 \\ 0 & \text{otherwise} \end{cases}$$

The wage ratio index is discretized by defining an interval including, but not centered upon, 0.5 in order to reflect the considerable skew in the distribution of the population across this index: 0.29-0.51 (see Figure 4d)[21]:

$$wagepwrlo = \begin{cases} 1 & \text{if } wagepwr < 0.29 \\ 0 & \text{otherwise} \end{cases}$$

$$wagepwrmed = \begin{cases} 1 & \text{if } (wagepwr \geq 0.29)\,\&\,(wagepwr < 0.51) \\ 0 & \text{otherwise} \end{cases}$$

$$wagepwrhi = \begin{cases} 1 & \text{if } wagepwr \geq 0.51 \\ 0 & \text{otherwise} \end{cases}$$

Each of these discretizations splits the population into groups containing roughly the same numbers of observations.

It is important to note that these four measures of power do indeed capture four very different aspects of power. The correlation matrix is presented below:

	age power	*education power*	*sex power*	*wage power*
age power	1.00			
education power	0.00	1.00		
sex power	−0.05	0.14	1.00	
wage power	0.02	0.19	−0.07	1.00

These correlations are very low. The highest correlation is only 0.19 and occurs between the education and wage power variables. These low correlations might be surprising at first glance, but do not forget that these are the correlations between the power indices which are not to be confused with the correlations between the underlying variables that do indeed present higher levels of correlation. For example, for the underlying education and wage variables, the correlation matrix is:

[21] The following more symetrical discretization was also tried, but it does not yield results that are statistically significant: age ratio index (0.49-0.51), sex ratio index (0.49-0.51), and wage ratio index (0.35-0.51). These results led to the third consideration presented above (i.e. the intervals need to be wide enough in order to identify groups with distinguishably different behaviours) and to the final discretization also presented above.

	education wife	*wage wife*	*education husband*	*wage husband*
education wife	1.00			
wage wife	0.35	1.00		
education husband	0.58	0.23	1.00	
wage husband	0.32	0.18	0.43	1.00

Here the lowest correlation is 0.18 and occurs between the wages within couples. All that this tells us is that although education and wages are correlated for couples, these correlations are not proportionate. Education within couples can be highly correlated and wages within couples can be highly correlated, but this does not necessarily imply that the difference in education within couples is correlated to the same degree with the difference in wages within couples. The advantage of this lack of correlation between the power indices is that multicollinearity will not be a problem for the precise estimation of the individual coefficients of these variables.

5.3 Individual and couple characteristics variables: g_h, g_w and b

Table 1 presents all of the control variables along with their definitions. In general, I use an expanded vector of control variables in the instrumenting regressions and a reduced vector of control variables in the subsequent regressions. The point of interest in the instrumenting regressions is the fitted values, and not the estimated coefficients. This implies that multicollinearity is not a problem and that I should include the most flexible specification possible.

For the individual characteristics γ_h and γ_w, I distinguish between personal and labor market characteristics. Personal characteristics are measured using age, education, race, religion, and health variables. Labor market characteristics are measured using work experience, occupation, and industry for the supply-side, and using the unemployment rate and geographic location for the demand-side. For the couple characteristics β, I use the number and age of children and other dependents, and the number of rooms and cars.

5.4 *Wage and income variables:* w_h, w_w, y_h, y_w and y

Table 1 also presents the wage and income variables along with their definitions. The variable *wealth* represents y as the net value of the couple's assets, excluding the value of home equity. The variable *other hh income* represents y_h and y_w as the sum of non-labor income for the couple and labor income for the spouse; this represents the income available to the couple if the individual were to choose not to do any market work. The variable *lwagehat* represents w_h and w_f as the log of estimated hourly wages earned by the individual. Since wages are not observed for those individuals who do not do any market work, I need to use the information contained in observed wages to calculate estimated wages for those individuals without observed wages. I do this using the two-stage Heckman procedure.[22]

The two-stage Heckman procedure consists in the estimation first of a participation equation using the entire sample, and then of a wage equation using the part of the sample for which wages are observed. The inverse Mill's ratio is calculated using the results from the first stage and is included as a regressor in the second stage to correct for sample selection bias. The selection regression includes all variables that could potentially contribute to explaining why an individual chooses to do or not to do any market work. These variables include personal and spousal characteristics (age, education, race, religion, health), couple characteristics (number and age of children and other dependants, other household income), and labor market conditions (unemployment rate). The wage regression includes all variables that could potentially affect the wage that an individual earns on the labor market. These variables include personal characteristics (age, education, race, health, learning disability), couple characteristics (number and age of children), and labor market considerations (work experience, occupation, industry, geographic location).

The results of the two-stage Heckman procedure are presented in Table 2. The only observation of particular interest is that whereas the coefficient of the sample selection correction term is significantly different from zero in the wife's wage equation, it is not in the husband's wage equation. This result confirms the accepted wisdom in the labor market

[22] I could alternatively restrict the sample to couples in which both spouses earn wages, but since I would like to understand the mechanisms underlying the different time allocation outcomes, I cannot just eliminate one particularly interesting time allocation outcome from the sample under study.

literature and just reflects the greater selection operating upon the labor market for women than for men.

5.5 Other control variable

In order to control for any systematic misreporting of the spouses' behavior, I include a dummy that is equal to one if the respondent to the 2001 PSID interview was the wife.

6. Empirical findings and Discussion

Estimation of the time allocation model is carried out in two stages. In order to account for the survey design of the data, observations are weighted using sampling weights and standard errors are corrected for clustering across strata and sampling units. This is achieved using the *svy* commands in STATA and defining *pweights*, *strata* and *psu*.

In the first stage, I instrument the four endogenous time variables and generate fitted values (see Table 3). In the second stage, I estimate the set of four simultaneous equations using an equation-by-equation method. I take this approach for two reasons. First, when using the alternate systems method, any specification error made in one equation contaminates the results for all of the other equations. Given the noisy data that I am working with, I want to minimize all other possible sources of error. Second, the fact that the dependant variables are continuous, binary, and ordinal, implies that classical systems methods are not applicable.

The housework equations are estimated using OLS, and the market work equations are estimated using ordered probit. This means that the marginal effects of the independent variables upon housework hours remain constant across observations, and therefore are straightforward to summarize. This also means that the marginal effects of the independent variables upon market work hours vary across observations, and therefore need to be evaluated for a representative couple. So, I need to identify the median couple, as defined in terms of the four time allocation variables.

Defining the median couple using the median values of the four time variables does not necessarily give us what we need. First, the wife with the median housework hours is not necessarily the wife with the median market work hours. Second, even if this did indeed happen to be the case, the same would have to hold true for the husband, and these median spouses would have to be married to one another. So, once again I need to locate the center of a four-dimensional cloud of observations. I define the median couple by calculating the median values of the four time variables for the 118 couples of the main population associated with a robust mahalanobis distance of less than one. The median couple has a wife who works 40 hours out of the home and 15 hours in the home, and a husband who works 48 hours out of the home and 6 hours in the home.

6.1 Time allocation process

Three aspects of the results on the household time allocation process stand out. The first aspect is the symmetry between the husband and the wife in the magnitudes of the estimated coefficients in their market work and housework equations. The second aspect is the asymmetry in the statistical significance of these coefficients. The third aspect is the statistical significance of *respondent* only in the husband's housework regression. I will discuss each of these results in turn.

6.1.1 Housework behavior

The estimation results of the housework equations for the husband and wife are summarized in the first two columns of Table 4 (see Table 5 for the full results). The estimated coefficients are expressed in minutes.

- The more market work one does, the less housework one does. An hour more of market work means 8 minutes less housework for the wife and 5 minutes less housework for the husband.

- The more market work one does, the more housework the spouse does. An hour more of the spouse's market work means 4 minutes more housework for both the husband and the wife alike.

- The more housework the spouse does, the more housework one does. An hour more of the spouse's housework means 7 minutes more housework for the wife and 12 minutes more for the husband.

These results attest to a striking symmetry in the way both spouses adjust their housework time in reaction to their own market work decisions and the other's housework and market work decisions. The asymmetry between the spouses' housework behavior lies in the statistical significance of one spouse's reaction to the other spouse's housework. Whereas the 7 minutes for the wife is not statistically different from zero, the 12 minutes for the husband is statistically different from zero.

6.1.2 Labor market behavior

The estimation results of the market work equations for the husband and wife are summarized in the last two columns of Table 4 (see Table 5 for the full results). The estimated coefficients are expressed in probabilities.

- The more housework one does, the higher the probability that one works shorter hours and the lower the probability that one works longer hours. An hour more of housework for the wife means 3% more chance that the wife does not work, 4% more chance that she works part-time, 1% less chance that she works fulltime and 6% less chance that she works overtime. An hour more of housework for the husband means 1% more chance that the husband works fulltime and 1% less chance that he works overtime.

- The more housework the spouse does, the lower the probability that one works shorter hours and the higher the probability that one works longer hours. An hour more of housework for the husband means 1% less chance that the wife does not work, 2% less chance that she works part-time, 1% more chance that she works fulltime and 2% more chance that she works overtime. An hour more of housework for the wife means 1% less chance that the husband works fulltime and 1% more chance that he works overtime.

- The more market work the wife does, the higher the probability (+1%) that the husband works shorter hours and the lower the probability (-1%) that he works longer hours. The husband's market work has no effect upon the wife's market work.

These results also reveal some symmetry in the way both spouses adjust their market work time in reaction to their own housework decisions and the other's housework and market work decisions. The asymmetry between the spouses' market work behavior lies in the statistical significance of their reactions to the housework variables and to the market work variables. While only the housework variables are significantly taken into account by the wife in her market work decision, only the market work variable is significantly taken into account by the husband in his market work decision.

6.1.3 Time allocation behavior

Putting the results from the market work and the housework equations together, we can draw the following profiles of the wife's and the husband's time allocation processes. The husband first decides upon his market work, taking his wife's market work decision into account, but only marginally. He then decides upon his housework taking everything (i.e. his and her market work and her housework) into account. So, the husband's behavior confirms the sociological approach to time allocation in which the market work decision is causally prior to the housework decision. The wife decides upon her market work taking housework into account, and decides upon her housework taking market work into account. So, the wife's behavior confirms the economic approach to time allocation in which market work and housework are simultaneously determined.

The last thing that needs to be discussed in relation to the basic time allocation model is the statistical significance of *respondent* only in the husband's housework regression. This result means that on average couples agreed upon the magnitude of each spouse's contributions in the two spheres, except for the husband's contribution to housework. When wives report their husbands' housework hours, they report 154 minutes less on average than when husbands' report their own housework hours. The question that naturally arises is whether it is the wives who are underreporting their husbands' housework hours or the other way around. Achen and Stafford (2005) study this issue but the results are inconclusive. The relevance of this result for this paper is that it very clearly indicates that the magnitude of the husband's contribution to housework is a source of disagreement within couples.

Can we go so far as to say that the wife's perception of her husband's contribution to the housework constrains her market work decision? Well, the results show that the more housework the husband does, the lower the chance that the wife works shorter hours and the higher the chance that the wife works longer hours. But if the husband does more housework then the wife will also do more housework (husband's and wife's housework are complements!) and this extra housework for the wife works in the opposite direction, increasing the chance that the wife works shorter hours and decreasing the chance that the wife works longer hours. The net impact of the husband's increased contribution to housework upon the wife's market work remains positive, but small.

6.2 Power variables

Once again the results highlight both symmetries and asymmetries between the spouses. The results suggest that spouses share a similar vision of the ideal balance between the home and the market, but because their starting points are so different, each spouse needs to move in opposite directions to get there. Not only do husbands and wives differ in how they use their power, they also differ in the sources from which they derive power. Wives derive power from education and wages and use that power to exchange housework hours for market work hours. Husbands derive power from sex ratios and use that power to exchange market work hours for housework hours. It would seem that the ideal place to be lies somewhere in between the wife's and the husband's current positions.

6.2.1 Housework behavior

The estimation results for the power variables in the housework equations are summarized in the first two columns of Table 4 (see Table 5 for the full results). I have three comments to make on these results. First, the estimated coefficients for the husbands and wives are of opposing signs for all but one of the power variables. Note that only 3 of these 16 estimated coefficients are significantly different from zero; nevertheless, the fact remains that on average husbands and wives use their power to move in opposite directions.

Second, in the regression for wives, the only power variable with an estimated coefficient that is significantly different from zero is the dummy for low wage power. Wives who earn wages that are lower than those earned by their husbands have to contribute 99 more minutes to housework than the same wives married to husbands with lower wages. So, wives use wage power to buy less hours of housework.

Third, in the regression for husbands, two of the power variables have estimated coefficients that are significantly different from zero, and both coefficients tell the same story that men use wage and remarriage power to buy more hours of housework. Low wage power is associated with 35 minutes less of housework and high remarriage power with 78 more minutes of housework. One possible explanation for this potentially perplexing result is that husbands and wives interpret the housework question differently, husbands including at least some childcare hours that wives do not include. This explanation provides the additional advantage of accounting for the 154 minutes difference in the husbands' and wives' reports of husbands' housework hours. In this case, the results would be indicating that men use their power to buy more time at home caring for their children.

6.2.2 Labor market behavior

The estimation results for the power variables in the market work equations are summarized in the last two columns of Table 4 (see Table 5 for the full results). In the regression for wives, two of the power variables have estimated coefficients that are significantly different from zero, and both coefficients tell the same story that wives use education and wage power to buy more hours of market work. Low education and wage power increase the probabilities of doing no or part-time market work and decrease the probabilities of doing fulltime or overtime market work. Low education power shifts 7% from the probabilities of working fulltime or more to the probabilities of working part-time or less. Low wage power shifts 22% in the same direction. Since wages are correlated with education, I would expect these effects to be often cumulated. What is fundamental about this result is that power can alter the effective labor supply of wives. I return to the example mentioned in the introduction. A 'typical' wife with 4 years of college education and an average wage, who is living with a husband with less education and a lower wage, has 14% less chance of working part-time and 11% more chance of working full-time when compared to the same woman who is living with a man with more education and a higher wage. So this same wife will most probably work full-time when she is in the more powerful situation with respect to her spouse, and part-time when she is in the less powerful situation.

In the regression for husbands, the only power variable with an estimated coefficient that is significantly different from zero is the dummy for high remarriage power. An increased concentration of women in the population is accompanied by an increased probability of working shorter hours. High remarriage power shifts 10% from the probability of working overtime to the probability of working fulltime. This result means that a husband who is living in a state where 50% of the population in his age-bracket is female has 10% less chance of working 40-45 hours a week and 10% more chance of working more than 45 hours a week, when compared to the same man living in a state where 51% of the population in his age-bracket is female. This result suggests that power can alter the effective labor supply of husbands. A husband will most probably have a ‘small’ full-time job when he is in the more powerful situation with respect to his partner, and a ‘big’ full-time job when he is in the less powerful situation.

7. Concluding remarks

Why are women less active than men on the labor market? Or alternatively, why are men more active than women on the labor market? Is it nature or nurture, choice or constraint? My results show that when wives benefit from relatively more power, defined in education and wage terms, they use this power to exchange housework hours for market work hours. (This result is compatible with the specialization and trade story.) When men benefit from relatively more power, defined in sex ratio terms, they use this power to exchange market work hours for housework hours. (This result is not compatible with the specialization and trade story.) So, the answer to the question would have to be « nurture » or « constraint ». Husbands and wives that are trapped in less powerful positions are constrained to choose sub-optimal levels of labor force participation. These results highlight two issues.

First, it would seem that men and women are more similar in their preferences on the work-family trade-off than the labor market statistics reveal. Women do less market work and more housework than husbands, but they would like to do more market work and less housework. Men do more market work and less housework than women, but they would like to do less market work and more housework. Men and women seem to agree that the ideal place to be is somewhere in between.

Second, the fact that what constitutes power is different for different people is fundamental to fully understanding the outcomes of intra-spousal bargaining. Indeed, both spouses of the same couple could simultaneously each feel less powerful than the other, the wife because she has less years of education and earns a lower wage than her husband, the husband because he faces a remarriage market with a relatively low concentration of women. This would generate a sub-optimal time allocation outcome for both of them, the wife working less than she would like to, and the husband working more than he would like to. This multi-dimensional aspect of the intra-spousal distribution of power appears to be a fruitful avenue for future research.

To conclude, this paper presents some evidence that cannot be reconciled with the classical specialization and trade story. The power story provides an alternative explanation of the data that poses some very tough social questions. Why do women work the hours that they do? Why do men work the hours that they do? Because they want to or because they have to? Depending on the answers to these questions, there could be profound implications for public policy. The riddle is far from resolved. Despite much debate in the press, less has taken place in the economics literature. Such a passionate debate would benefit from more econometric rigor. I hope to have contributed in such a manner.

8. References

Achen, A.C., and F.P. Stafford, 2005, Data quality of housework hours in the panel study of income dynamics: who really does the dishes?, Mimeo.

Anxo, D., and P. Carlin, 2004, Intra-family time allocation to housework – French evidence, electronic International Journal of Time Use Research 1, 14-36.

Bauchner, E., 2003, Mothering matters: the “opt-out” revolution hasn’t begun, Ithaca Journal, November 18.

Belkin, L., 2003, The opt-out revolution, The New York Times Magazine October 26.

Bianchi, S.M., M.A. Milchie, L.C. Sayer, and J.P. Robinson, 2000, Is anyone doing housework? Trends in the gender division of household labor, Social Forces 79, 191-222.

Bruegel, I., 1996, Whose myths are they anyway? A comment, British Journal of Sociology 47, 175-177.

Burk, M., P. Gillette, and D. Rhode, 2005, The opt-out revolution, The Stanford Center of Ethics Newsletter, June.

Castleman, T., and R. Reed, 2003, ‘One day to have my own family…’: ideas about a successful life from early professionals in a longitudinal study, Australian Institute for Family Studies Conference, October 12-14, Melbourne.

Chiappori, P.A., B. Fortin, and G. Lacroix, 2002, Marriage market, divorce legislation, and household labor supply, The Journal of Political Economy 110, 37-72.

Douglas, S.J., 2003, Mommas in the marketplace, In These Times, November 17.

Duesenberry, J.S., 1960, Comment on Becker’s ‘An economic analysis of fertility’ In National Bureau Committee for Economic Research, Demographic and Economic Change in Developed Countries, Princeton University Press, Princeton, N.J.

Ghysels, J., 2004, Work, family and childcare: an empirical analysis of European households, Edward Elgar, Cheltenham.

Ginn, J., S. Arber, J. Brannen, A. Dale, S. Dex, P. Elias, P. Moss, J. Pahl, C. Roberts, and J. Rubery, 1996, Feminist fallacies: a reply to Hakim on women’s employment, British Journal of Sociology 47, 167-174.

Godwin, D.D., 1991, Spouses' time allocation to household work: a review and critique, Lifestyles: Family and Economic Issues 12, 253-294.

Goldin, C., 2006, The quiet revolution that transformed women's employment, education and family, AEA Papers and Proceedings 96, 1-21.

Gray, J.S., 1998, Divorce-law changes, household bargaining, and married women's labor supply, The American Economic Review 88, 628-642.

Grossbard-Shechtman, S., and C. Amuedo-Dorantes, 2003, When are husbands more likely to pay the bills? Marriage market effects on the aggregate labor supply of married women, Mimeo.

Hakim, C., 1995, Five feminist myths about women's employment, British Journal of Sociology 46, 429-455.

Hakim, C., 2000, Work-lifestyle choices in the 21st century, Oxford University Press, Oxford.

Hochschild, A., and A. Machung, 1989, The Second Shift, Penguin, New York.

Juster, F.T., H. Ono, and F.P. Stafford, An assessment of alternative measures of time use, Sociological Methodology 33, 19-54.

Kan, M.Y., 2005, Work orientation and wives' employment careers: an evaluation of Hakim's preference theory, ISER Working Paper 2005-27, University of Essex, Colchester.

Kan, M.Y., 2006, Measuring housework participation: the gap between "stylized" questionnaire estimates and diary-based estimates, ISER Working Paper 2006-11, University of Essex, Colchester.

Lundberg, S., 1988, Labor supply of husbands and wives: a simultaneous equations approach, The Review of Economics and Statistics 70, 224-235.

Marini, M.M., and B.A. Shelton, 1993, Measuring household work: recent experience in the United States, Social Science Research 22, 361-382.

Martinez-Granado, M., 2004, Testing labour supply and hours constraints, Labour Economics 12, 321-343.

Mattila-Wiro, P., 1999, Economic theories of the household: a critical review, UNU/WIDER Working Paper 159.

McRae, S., 2003a, Constraints and choices in mothers' employment careers: a consideration of Hakim's preference theory, British Journal of Sociology 53, 317-338.

McRae, S., 2003b, Choice and constraints in mothers' employment careers: McRae replies to Hakim, British Journal of Sociology 54, 585-592.

Pollack, R.A., 1969, Conditional demand functions and consumption theory, The Quarterly Journal of Economics 83, 60-78.

Pollack, R.A., 1978, Endogenous tastes in demand and welfare analysis, The American Economic Review 68, 374-379.

Pollack, R.A., 2003, Gary Becker's contributions to family and household economics, Review of Economics of the Household 1, 111-141.

Pollack, R.A., 2005, Bargaining power in marriage: earnings, wage rates and household production, NBER Working Paper 11239.

Pollitt, K., 2003, There they go again, The Nation , November 17.

Procter, I., and M. Padfield, Work orientations and women's work: a critique of Hakim's theory of the heterogeneity of women, Gender, Work and Organization 6, 152-162.

Rappoport, R., L. Bailyn, J. K. Fletcher, and B. H. Pruitt, 2002, Beyond work-family balance: Advancing gender equity and workplace performance, Jossey-Bass, San Francisco, CA.

Robinson, J.P., 1985, The validity and reliability of diaries versus alternative time use measures In Juster, F.T., and F.P. Stafford (Eds.), Time, Goods and Well-being, University of Michigan, Ann Arbor, MI.

Rousseeuw, P.J., and B.C. van Zomeren, 1990, Unmasking multivariate outliers and leverage points, Journal of the American Statistical Association 85, 633-639.

Stigler, G.J., and G.S. Becker, 1977, De gustibus non est disputandum, The American Economic Review 67, 76-90.

The Economist, 2006, Special Report: The American Economy – Slow Road Ahead, The Economist, October 28 – November 3, 85-86.

U.S. Glass Ceiling Commission, 1995, Good for business: making full use of the nation’s human capital, U.S. Government Printing Office, Washington, D.C.

Walsh, J., 2003, Clueless in Manhattan, Salon.com, October 27.

Weber, M., 1962, Basic Concepts in Sociology, Citadel, New York.

Young, C., 2004, Opting out: the press discovers the mommy wars, again, Reason.com, June.

Figure 1: Husbands' and wives' weekly housework and market work hours

(Note: Histograms are constructed using 2001 PSID data for cohabiting couples.)

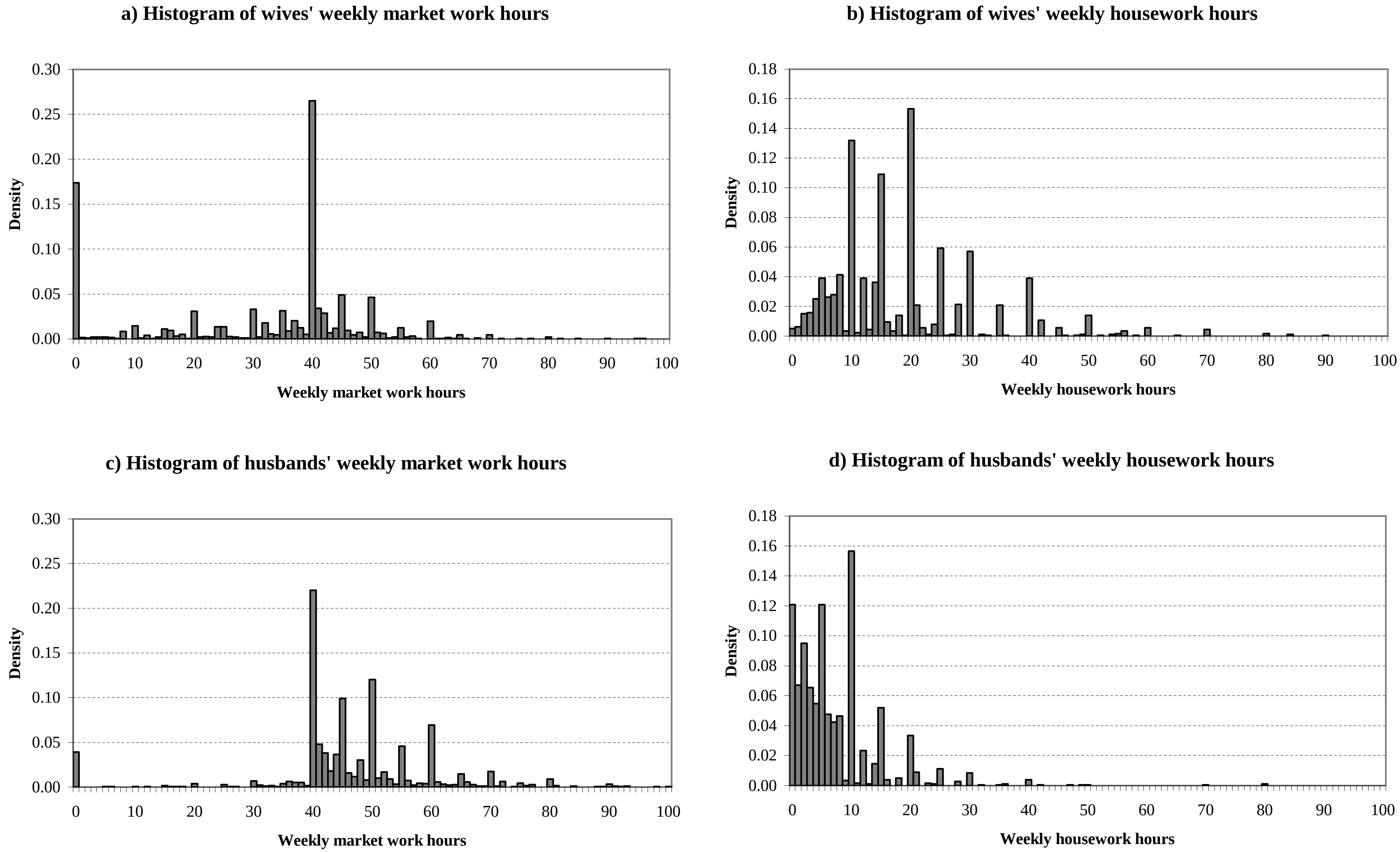
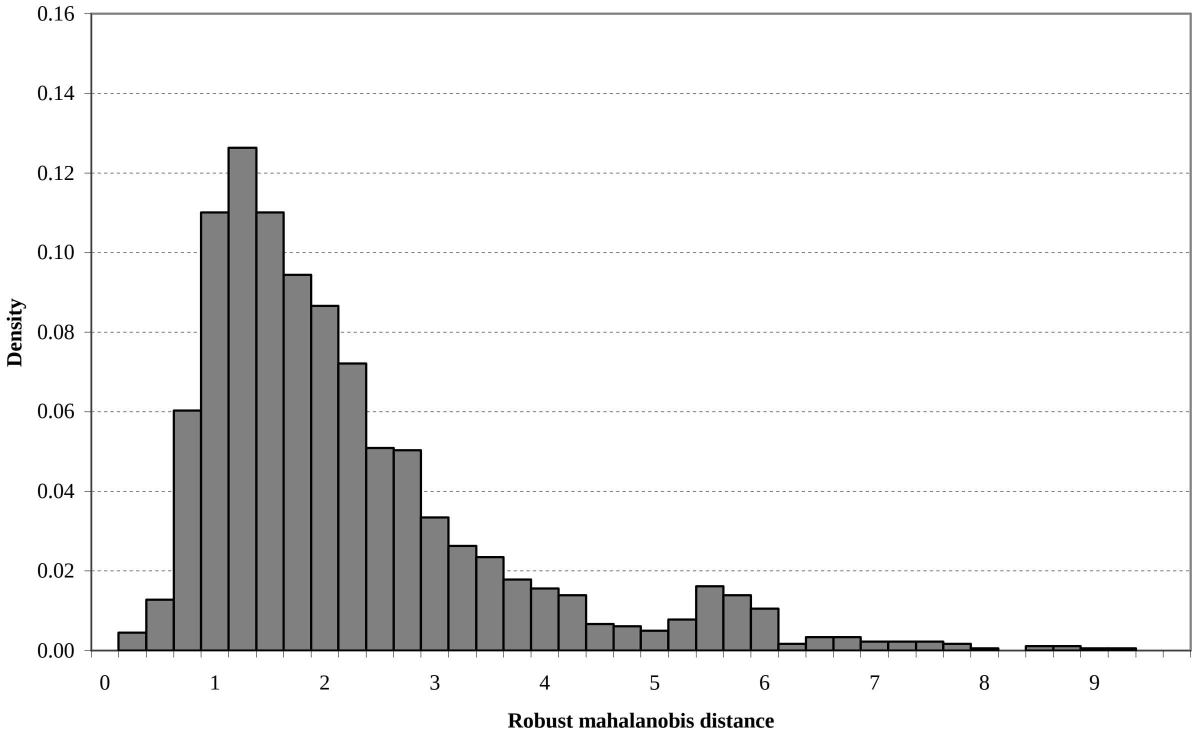


Figure 2: Distribution of robust mahalanobis distances for entire population

(Notes: Calculations of robust distances are based on 2001 PSID data for cohabiting couples. The presence of a sub-population composed of couples in which the husbands do not do any market work is revealed by the cluster of observations located at RMD=5.5 in graph a and RMD=22 in graph b. The dependent variables are husbands' and wives' time spent on market and housework per week. All variables also include each spouses' age, education, work experience, full-time work experience, work weeks, and wage, and the couples' number of children.)

a) Robust distances calculated in function of dependent variables



b) Robust distances calculated in function of all variables

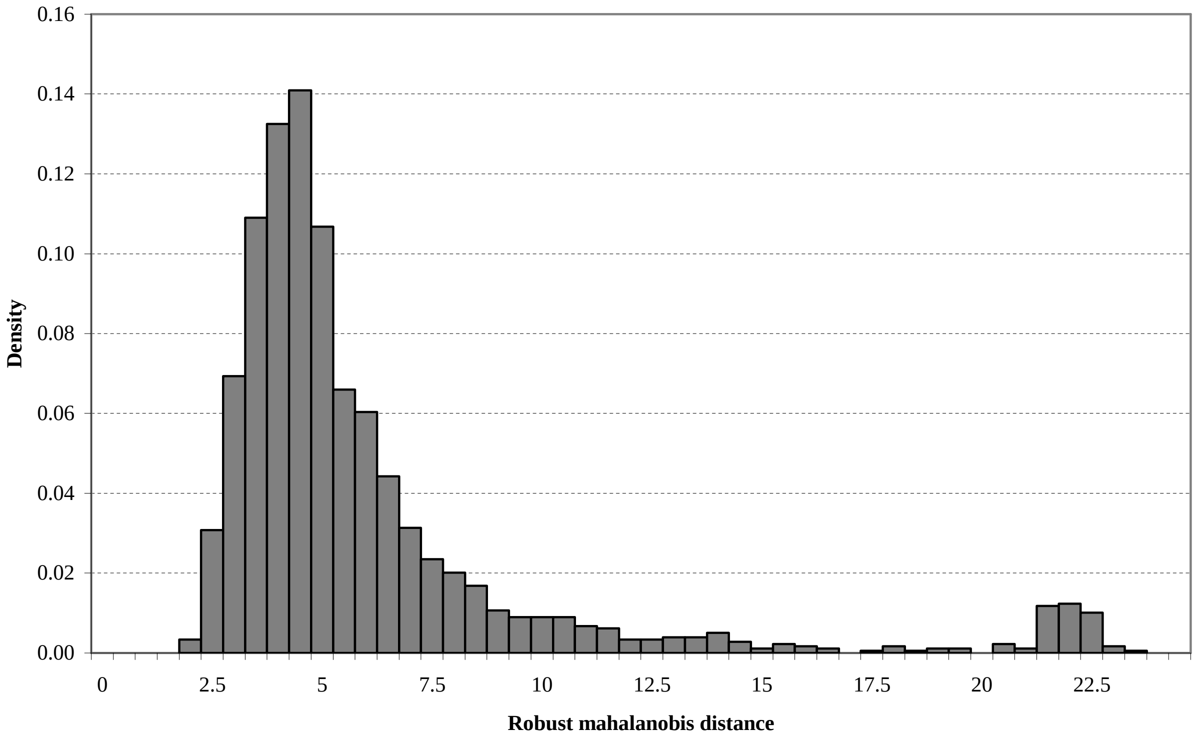
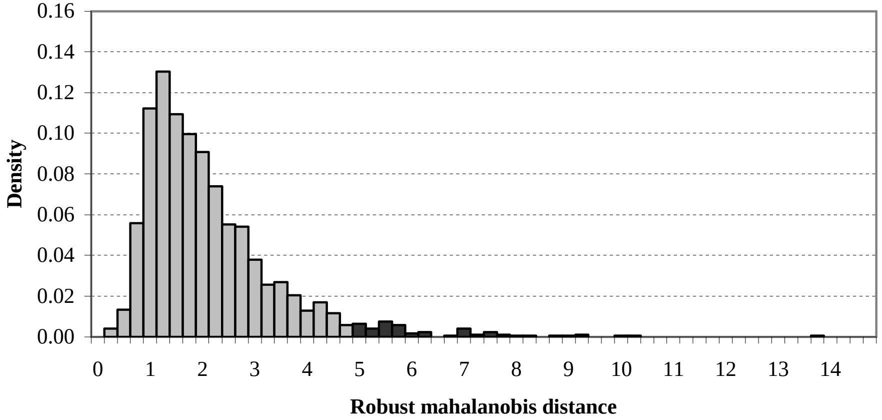


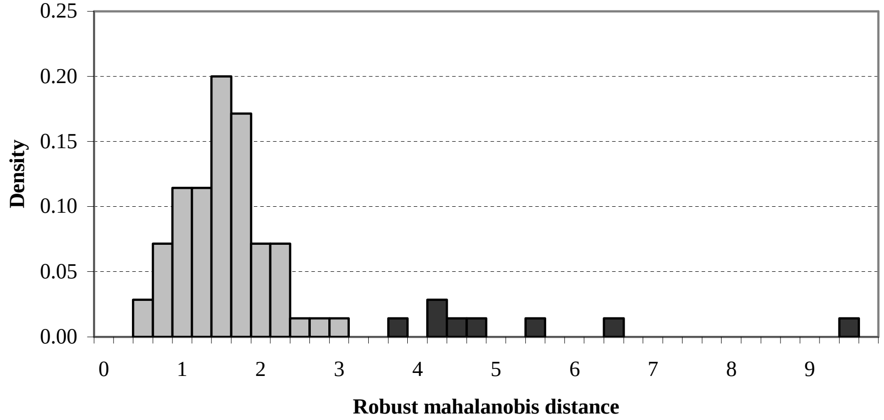
Figure 3: Distribution of mahalanobis robust distances for the main and the sub-populations

(Notes: Outlying observations that were eliminated from the sample are marked in black. Calculations of robust distances are based on 2001 PSID data for cohabiting couples. The main population is composed of couples in which the husband works a positive number of hours on the labor market weekly. The sub-population is composed of couples in which the husband does not do any market work. The dependent variables are husbands' and wives' time spent on market and housework per week. All variables also include each spouses' age, education, work experience, full-time work experience, work weeks, and wage, and the couples' number of children.)

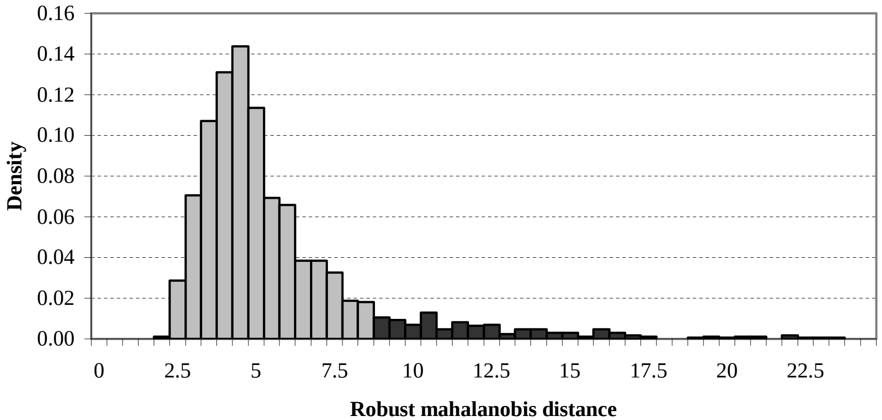
a) Robust distances for main population
(calculations based on dependent variables)



c) Robust distances for sub-population
(calculations based on dependent variables)



b) Robust distances for main population
(calculations based on all variables)



d) Robust distances for subpopulation
(calculations based on all variables)

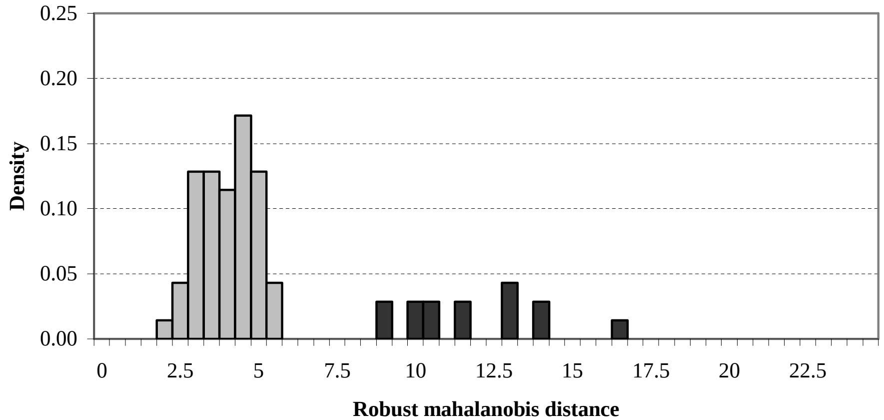
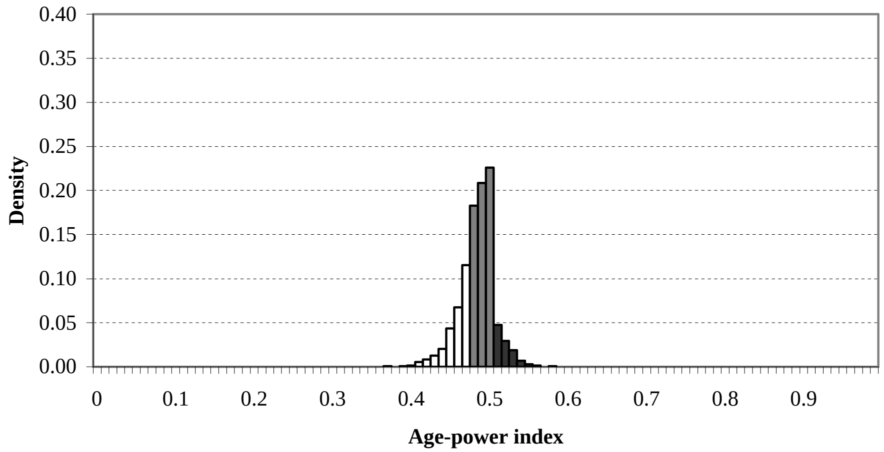


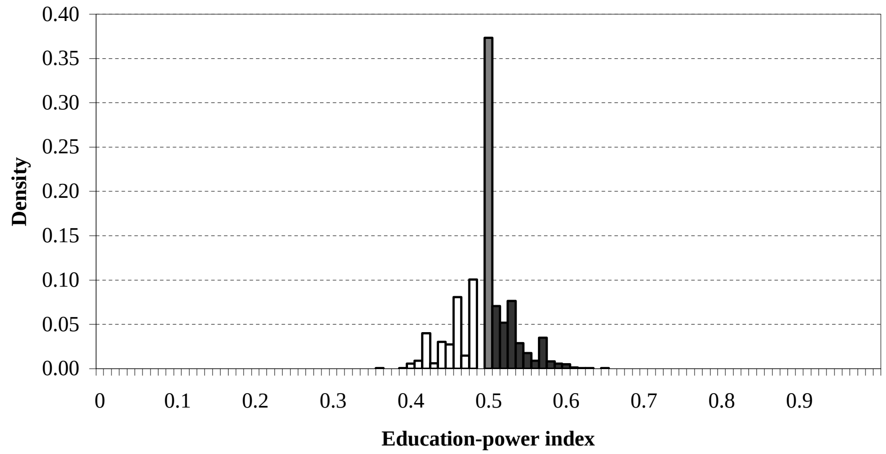
Figure 4: Power distribution within couples

(Notes: Histograms are constructed using 2001 PSID data for cohabiting couples. Power is an index measuring the wife's contribution to the couple's resources in age, education, wages and sex ratios. This index varies from 0 to 1, 0 representing no power for the wife, 1 total power for the wife, and 0.5 equality between the two spouses. The discretization used in the empirical analysis is indicated by the colors of the bars, white indicating more power for the huband, grey equal power between the spouses, and black more power for the wife.)

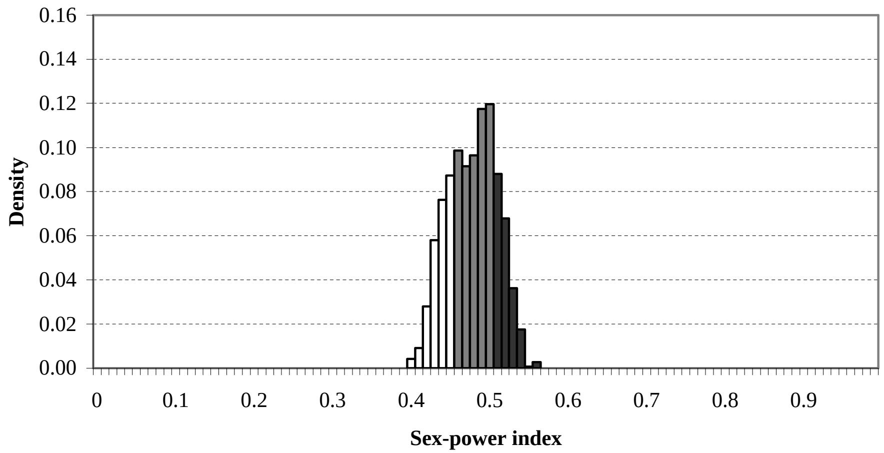
a) Age-power distribution within couples



b) Education-power distribution within couples



c) Sex-power distribution within couples



d) Wage-power distribution within couples

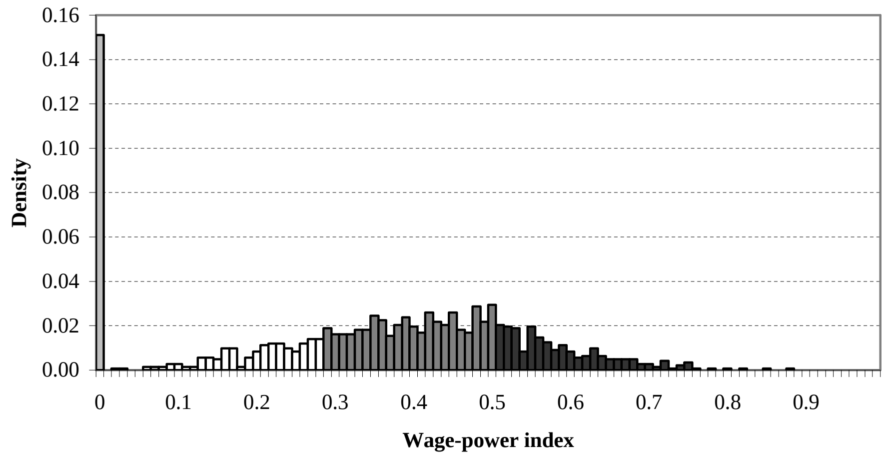


Table 1: Definitions of control variables

Variable	**Definition in terms of PSID references**		**Definition in general terms**
	Husband	Wife	
Personal characteristics of the individual			
Age variables			
age	ER17013-mean(ER17013)	ER17015-mean(ER17015)	Age in years, centered upon mean
age2	age^2	age^2	Age, centered and squared
age3	age^3	age^3	Age, centered and cubed
Education variables			
education	UPEDU01H-mean(UPEDU01H)	UPEDU01W-mean(UPEDU01W)	Education in years, centered upon mean
education2	$education^2$	$education^2$	Education, centered and squared
education3	$education^3$	$education^3$	Education, centered and cubed
edudummy2	UPEDU01H=12-15	UPEDU01W=12-15	Dummy equal to one for individuals with 12-15 years of education
edudummy3	UPEDU01H=16-17	UPEDU01W=16-17	Dummy equal to one for individuals with 16-17 years of education
Age - Education interaction terms			
*age*edu*	age*education	age*education	Age multiplied by years of education
*age*edu2*	age*education2	age*education2	Age multiplied by education squared
*age2*edu*	age2*education	age2*education	Age squared multiplied by education
*age*edu3*	age*education3	age*education3	Age multiplied by education cubed
*age3*edu*	age3*education	age3*education	Age cubed multiplied by education
Race dummies			
race1	ER19989=1	ER19897=1	The base category is composed of individuals of white race.
race2	ER19989=2	ER19897=2	Dummy equal to one for individual of black race
race3	ER19989=3-7	ER19897=3-7	Dummy equal to one for individual of other races

Continued on the next page

Table 1 (continued)

Variable	Definition in terms of PSID references		Definition in general terms
	Husband	Wife	
Personal characteristics of the individual (continued)			
Religion dummies	ER20038 updated using ER15977, ER11895, ER9099, ER6853, ER3983 & V23315.	ER19945 updated using ER15884, ER11807, ER9029, ER6783, ER3913 & V23242.	
religion1	ER20038updated=3-9	ER19945updated=3-9	The base category is composed of individuals of protestant religion.
religion2	ER20038updated=1, 13	ER19945updated=1, 13	Dummy equal to one for individuals of catholic religion
religion3	ER20038updated=11-12, 14-25	ER19945updated=11-12, 14-25	Dummy equal to one for individuals of restorationist religions
religion4	ER20038updated=2	ER19945updated=2	Dummy equal to one for individuals of jewish religion
religion5	ER20038updated=10, 97	ER19945updated=10, 97	Dummy equal to one for individuals of other religions
religion6	ER20038updated=0	ER19945updated=0	Dummy equal to one for atheists or agnostics
Other variables			
bad health	ER19612=4-5	ER19720=4-5	Dummy equal to one for individuals claiming to be in bad health
learning disability	ER19683=1	ER19791=1	Dummy equal to one for individuals claiming to have a learning disability
Labor market characteristics of the individual			
Occupation dummies	HDOCC01 updated using: HDOCC99, HDOCC97, HDOCC96, HDOCC95, HDOCC94 & V22456.	WFOCC01 updated using: WFOCC99, WFOCC97, WFOCC96, WFOCC95, WFOCC94 & V22809.	
occupation1	HDOCC01updated=1-195	WFOCC01updated=1-195	The base category is composed of professional, technical and kindred workers.
occupation2	HDOCC01updated=201-245	WFOCC01updated=201-245	Dummy equal to one for managers and administrators, except farm.
occupation3	HDOCC01updated=401-600	WFOCC01updated=301-395	Wives: Dummy equal to one for clerical and kindred workers. Husbands: Dummy equal to one for craftsmen and kindred workers.
occupation4	HDOCC01updated=601-695	WFOCC01updated=901-965	Wives: Dummy equal to one for service workers, except private household. Husbands: Dummy equal to one for operatives, except transport.
occupation5	HDOCC01updated=other values	WFOCC01updated=other values	Dummy equal to one for other occupations, including no occupation

Continued on the next page

Table 1 (continued)

Variable	**Definition in terms of PSID references**		**Definition in general terms**
	Husband	Wife	
Labor market characteristics of the individual (continued)			
Industry dummies	HDIND01 updated using: HDIND99, HDIND97, HDIND96, HDIND95, HDIND94 & V22457.	WFIND01 updated using: WFIND99, WFIND97, WFIND96, WFIND95, WFIND94 & V22810.	
industry1	HDIND01updated=107-398	WFIND01updated=107-398	The base category is composed of individuals with job in manufacturing sector.
industry2	HDIND01updated=407-479	WFIND01updated=407-479	Dummy equal to one for individual with job in transportation, communications an other public utilities sectors.
industry3	HDIND01updated=507-698	WFIND01updated=507-698	Dummy equal to one for individual with job in wholesale and retail trade sectors.
industry4	HDIND01updated=828-897	WFIND01updated=828-897	Dummy equal to one for individual with job in professional and related services.
industry5	HDIND01updated=other values	WFIND01updated=other values	Dummy equal to one for individual with job in other industries, including no industry.
Other variables			
work experience	ER20040 updated using: ER15979, ER11897, ER9101, ER6855, ER3985 & V23316.	ER19947 updated using: ER15886, ER11809, ER9031, ER6785, ER3915 & V23243.	Number of years worked since 18 years of age.
ft work experience	ER20041 updated using: ER15980, ER11898, ER9102, ER6856, ER3986 & V23317.	ER19948 updated using: ER15887, ER11810, ER9032, ER6786, ER3916 & V23244.	Number of years worked full-time since 18 years of age.
work weeks	HDWKS01	WFWKS01	Number of weeks worked during 2001
weeksdummy2	HDWKS03=31-47	WFWKS01=31-47	Dummy equal to one for individual who worked between 31 and 48 weeks in 200
weeksdummy3	HDWKS04=48-52	WFWKS01=48-52	Dummy equal to one for individual who worked more than 48 weeks in 2001.
log wage	log(HDWGE01)	log(WFWGE01)	Log of hourly wages
non-labor income	ER18634 + ER18650 + ER18666 + ER18682 + ER18847 + ER18863	ER18966 + ER18982 + ER18998 + ER19014 + ER19063 + ER19143	Non-labor income for the individual in 2001
other hh income	couple's non-labor income + WFEARN01	couple's non-labor income + HDEARN01	Sum of non-labor income for the couple and labor income for the spouse in 2001

Continued on the next page

Table 1 (continued)

Variable	**Definition in terms of PSID references** Couple	**Definition in general terms**
Other variables (continued)		
beale1 **-** *beale10*	ER20377=1 - ER20377=10	10 dummies for the ten values taken by the Beale-Ross rural-urban continuum code for 2001 residence, ranging from central counties of metropolitan areas of 1 million population or more (beale1=1) to completely rural, not adjacent to a metropolitan area (beale=10)
unemployment rate	Census data	Rate of unemployment for state of residence in 2001
Characteristics of the couple		
Children dummies		The base category is composed of couples with no children.
child1	ER17016=1	Dummy equal to one for couples with one child
child2	ER17016=2	Dummy equal to one for couple with two children
child3	ER17016>2	Dummy equal to one for couples with more than than two children
youngchild	ER17017<6	Dummy equal to one for couples with a child that is younger than 6 years old
Other variables		
add hh members	ER17018	Number of additional household members not belonging to the immediate family
rooms	ER17040	Number of rooms for the household, not including bathrooms
cars	ER17111	Number of cars owned or leased by the household
wealth	S516	Net value of couple's assets, excluding value of home equity
respondent	ER17019	Dummy equal to one if the respondent to the 2001 PSID interview is the wife

Table 2: Heckman two-stage procedure

	Participation equations		Wage equations	
	Wife	Husband	Wife	Husband
Personal characteristics of the individual				
age	0.013	0.016	-0.001	0.010
	(0.73)	(0.52)	(-0.15)	(2.39)
age2	-0.001	0.000	0.000	-0.001
	(-1.24)	(0.27)	(-2.75)	(-2.84)
age3	0.000	0.000	0.000	0.000
	(-1.36)	(-0.85)	(0.92)	(2.57)
education	0.048	0.173	0.070	0.083
	(0.78)	(1.79)	(4.28)	(5.49)
education2	-0.010	-0.037	0.004	0.002
	(-0.87)	(-1.15)	(0.84)	(0.51)
education3	0.003	-0.009	-0.001	-0.001
	(0.56)	(-1.15)	(-0.77)	(-1.05)
*age*edu*	0.019	0.012	-0.002	-0.004
	(2.67)	(1.11)	(-0.61)	(-1.92)
*age*edu2*	-0.001	-0.005	0.000	0.000
	(-0.84)	(-1.38)	(-0.78)	(0.64)
*age2*edu*	0.000	0.001	0.000	0.000
	(0.07)	(1.40)	(-0.46)	(0.55)
*age*edu3*	-0.001	-0.001	0.000	0.000
	(-1.20)	(-1.48)	(1.22)	(1.17)
*age3*edu*	0.000	0.000	0.000	0.000
	(-2.91)	(-0.22)	(0.18)	(1.00)
race2	-0.566	-0.311	-0.084	-0.197
	(-1.37)	(-1.14)	(-1.66)	(-4.20)
race3	0.241	-0.088	0.068	-0.235
	(0.92)	(-0.24)	(0.56)	(-2.44)
learning disability	–	–	0.068	0.030
			(1.52)	(0.88)
bad health	-0.791	-1.165	0.074	-0.225
	(-5.78)	(-4.90)	(1.18)	(-3.57)
religion2	-0.021	0.088	–	–
	(-0.15)	(0.31)		
religion3	-0.023	0.285	–	–
	(-0.10)	(0.42)		
religion4	0.190	5.288	–	–
	(0.67)	(11.53)		
religion5	-0.203	-0.049	–	–
	(-0.64)	(-0.11)		
religion6	0.319	0.486	–	–
	(1.72)	(1.50)		

Continued on the next page

Table 2 (continued)

	Participation equations		Wage equations	
	Wife	Husband	Wife	Husband
Labor market characteristics of the individual				
occupation2	–	–	-0.082 (-1.65)	0.049 (1.13)
occupation3	–	–	-0.215 (-5.27)	-0.143 (-2.84)
occupation4	–	–	-0.361 (-5.01)	-0.156 (-2.41)
occupation5	–	–	-0.318 (-4.61)	-0.259 (-5.39)
industry2	–	–	0.009 (0.15)	-0.028 (-0.67)
industry3	–	–	-0.234 (-2.47)	-0.229 (-5.46)
industry4	–	–	-0.081 (-1.62)	-0.254 (-4.07)
industry5	–	–	-0.047 (-1.01)	-0.116 (-6.39)
beale2	–	–	0.041 (0.95)	0.037 (0.81)
beale3	–	–	-0.096 (-2.13)	-0.128 (-3.05)
beale4	–	–	-0.235 (-3.29)	-0.204 (-3.82)
beale5	–	–	-0.121 (-1.72)	-0.245 (-3.70)
beale6	–	–	-0.145 (-2.1)	-0.203 (-2.47)
beale7	–	–	-0.316 (-5.24)	-0.252 (-4.04)
beale8	–	–	-0.396 (-6.93)	-0.385 (-4.58)
beale9	–	–	-0.157 (-1.77)	-0.319 (-2.11)
beale10	–	–	-0.453 (-4.98)	-0.513 (-4.97)
work experience	–	–	0.009 (1.54)	-0.020 (-2.79)
ft work experience	–	–	0.008 (1.57)	0.004 (0.64)
unemployment rate	-0.473 (-0.11)	-0.156 (-0.02)	–	–

Continued on the next page

Table 2 (continued)

	Participation equations		Wage equations	
	Wife	Husband	Wife	Husband
Characteristics of the spouse				
race2	1.179	-0.225	–	–
	(3.01)	(-0.94)		
race3	0.047	-0.353	–	–
	(0.14)	(-1.03)		
bad health	0.035	0.051	–	–
	(0.19)	(0.18)		
religion2	0.069	-0.296	–	–
	(0.54)	(-1.15)		
religion3	0.063	-0.386	–	–
	(0.27)	(-0.63)		
religion4	0.405	4.646	–	–
	(1.88)	(10.16)		
religion5	-0.200	-0.085	–	–
	(-0.53)	(-0.19)		
religion6	-0.087	0.132	–	–
	(-0.70)	(0.39)		
Characteristics of the couple				
child1	-0.211	0.961	-0.016	0.039
	(-1.56)	(3.28)	(-0.36)	(1.12)
child2	-0.300	0.750	-0.046	0.033
	(-2.08)	(3.14)	(-1.06)	(0.81)
child3	-0.437	0.219	-0.108	0.068
	(-2.67)	(0.69)	(-1.58)	(1.51)
youngchild	-0.217	-0.029	0.186	0.057
	(-1.85)	(-0.10)	(3.56)	(1.74)
add hh members	-0.104	-0.071	–	–
	(-0.72)	(-0.18)		
other hh income	0.000	0.000	–	–
	(-1.75)	(-2.46)		
wealth	0.000	0.000	–	–
	(-0.34)	(-1.11)		
inverse mill's ratio	–	–	-1.098	0.228
			(-5.62)	(0.87)
constant	1.401	2.329	2.516	3.563
	(4.99)	(3.36)	(10.03)	(17.39)
(pseudo) R^2	0.088	0.298	0.377	0.389

Notes: The results are OLS estimations. Observations are weighted using sampling weights. Standard errors are adjusted for clustering across strata and sampling units. The t-statistics are in the brackets.

Table 3: Instrumenting regressions for house and market work variables

	Housework equations		Market work equations	
	Wife	Husband	Wife	Husband
Personal characteristics of the individual				
age	0.132	-0.083	-0.043	0.045
	(1.68)	(-1.72)	(-0.24)	(0.30)
age2	-0.001	-0.007	0.005	0.002
	(-0.23)	(-3.14)	(0.93)	(0.43)
age3	0.000	0.000	-0.001	0.000
	(1.01)	(-0.53)	(-1.87)	(0.33)
education	-0.755	0.215	0.483	0.906
	(-1.8)	(1.68)	(1.38)	(2.28)
education2	0.034	-0.026	0.017	0.019
	(0.34)	(-0.59)	(0.14)	(0.23)
education3	0.018	-0.006	-0.036	0.003
	(0.67)	(-0.55)	(-1.20)	(0.11)
*age*edu*	-0.045	-0.029	0.054	-0.074
	(-1.17)	(-1.5)	(0.97)	(-1.44)
*age*edu2*	0.000	0.010	0.018	0.000
	(0.01)	(2.23)	(1.58)	(0.01)
*age2*edu*	0.000	0.000	0.002	-0.001
	(0.05)	(-0.32)	(0.82)	(-0.52)
*age*edu3*	-0.001	0.002	0.009	0.003
	(-0.56)	(2.04)	(2.34)	(0.82)
*age3*edu*	0.000	0.000	-0.001	0.000
	(0.46)	(0.8)	(-3.35)	(1.51)
race2	-4.906	2.046	-1.703	0.131
	(-2.75)	(1.85)	(-0.50)	(0.05)
race3	-2.808	0.244	0.516	1.743
	(-1.8)	(0.49)	(0.28)	(0.72)
religion2	-0.261	-0.048	-0.710	0.364
	(-0.37)	(-0.17)	(-0.63)	(0.62)
religion3	-1.013	-0.058	0.405	-0.318
	(-0.85)	(-0.07)	(0.24)	(-0.16)
religion4	-1.045	-0.708	1.483	2.188
	(-0.68)	(-0.67)	(0.47)	(1.76)
religion5	0.289	-1.301	-0.995	-1.983
	(0.19)	(-1.48)	(-0.53)	(-0.7)
religion6	-1.960	0.312	0.525	-0.247
	(-2.07)	(0.52)	(0.36)	(-0.21)
Labor market characteristics of the individual				
occupation2	-0.700	-1.040	2.146	1.738
	(-0.82)	(-1.64)	(2.19)	(2.13)
occupation3	1.132	0.238	-1.774	1.231
	(1.5)	(0.35)	(-1.90)	(1.35)
occupation4	1.100	-0.691	-3.365	-1.391
	(0.87)	(-0.73)	(-2.20)	(-1.74)
occupation5	1.240	-0.106	-3.734	0.847
	(1.06)	(-0.15)	(-3.54)	(0.86)

Continued on the next page

Table 3 (continued)

	Housework equations		Market work equations	
	Wife	Husband	Wife	Husband
Labor market characteristics of the individual (continued)				
industry2	-1.337	-0.694	-0.751	2.384
	(-0.76)	(-1.21)	(-0.45)	(2.27)
industry3	3.650	-0.645	-1.411	1.596
	(4.21)	(-1.12)	(-1.25)	(1.70)
industry4	1.011	-0.369	-2.430	-0.296
	(0.92)	(-0.46)	(-2.48)	(-0.35)
industry5	1.999	-1.231	-2.959	0.811
	(2.11)	(-2.58)	(-3.68)	(1.12)
work weeks	-0.203	-0.057	0.608	0.843
	(-11.85)	(-3.46)	(28.22)	(30.53)
work experience	–	–	-0.472	-0.197
			(-4.13)	(-0.93)
ft work experience	–	–	0.842	0.302
			(8.23)	(1.60)
unemployment rate	-18.212	1.891	-33.481	-56.932
	(-0.76)	(0.09)	(-1.02)	(-2.00)
Characteristics of the spouse				
race2	1.739	-1.774	3.333	-0.860
	(0.91)	(-1.74)	(0.97)	(-0.30)
race3	0.064	0.484	0.485	0.181
	(0.05)	(0.59)	(0.31)	(0.08)
religion2	1.133	0.095	-0.620	0.108
	(1.67)	(0.24)	(-0.85)	(0.11)
religion3	2.304	-0.425	-1.662	1.146
	(1.22)	(-0.58)	(-0.85)	(0.67)
religion4	-2.434	-0.090	-1.943	-1.105
	(-1.37)	(-0.09)	(-0.74)	(-0.73)
religion5	-0.260	0.769	-0.545	0.351
	(-0.16)	(0.7)	(-0.34)	(0.19)
religion6	-0.610	0.131	-0.773	-2.084
	(-0.56)	(0.17)	(-0.63)	(-1.98)
occupation2	-0.508	0.789	0.950	0.673
	(-0.56)	(1.48)	(1.28)	(0.86)
occupation3	-0.712	0.828	1.674	0.469
	(-0.81)	(1.63)	(1.61)	(0.61)
occupation4	-4.933	0.405	2.765	-0.389
	(-3.17)	(0.92)	(1.55)	(-0.39)
occupation5	-1.163	0.715	0.683	0.641
	(-1.45)	(1.21)	(0.68)	(0.73)

Continued on the next page

Table 3 (continued)

	Housework equations		Market work equations	
	Wife	Husband	Wife	Husband
Characteristics of the spouse (continued)				
industry2	-1.206	-0.122	1.828	0.199
	(-1.22)	(-0.1)	(2.05)	(0.11)
industry3	-1.850	0.316	0.961	-0.918
	(-2.16)	(0.73)	(0.97)	(-0.75)
industry4	-1.660	-0.168	3.878	0.810
	(-1.76)	(-0.3)	(3.24)	(0.72)
industry5	-1.955	-0.154	0.523	0.106
	(-2.8)	(-0.28)	(0.68)	(0.09)
work weeks	0.106	0.017	-0.018	-0.064
	(3.11)	(1.71)	(-0.51)	(-4.14)
Characteristics of the couple				
child1	3.033	0.286	-0.551	0.834
	(3.91)	(0.52)	(-0.49)	(0.91)
child2	5.593	1.009	-1.568	1.389
	(6.75)	(1.67)	(-1.60)	(1.46)
child3	6.719	0.504	-1.732	0.572
	(5.37)	(0.67)	(-1.06)	(0.51)
youngchild	0.817	-0.160	-2.689	-1.084
	(1.31)	(-0.35)	(-2.83)	(-1.08)
add hh members	1.386	-0.593	-0.365	-0.332
	(1.18)	(-0.72)	(-0.31)	(-0.20)
rooms	-0.121	0.176	–	–
	(-0.87)	(1.87)		
cars	-0.283	0.202	–	–
	(-0.97)	(1.1)		
age power	-2.918	-2.588	-13.302	11.305
	(-0.15)	(-0.42)	(-0.91)	(0.88)
education power	26.302	5.309	9.673	9.918
	(2.39)	(1.38)	(1.00)	(1.02)
sex power	-23.857	-10.598	31.421	50.517
	(-1.42)	(-0.86)	(0.99)	(2.20)
respondent	-0.312	-2.592	-0.011	-0.433
	(-0.6)	(-9.98)	(-0.02)	(-0.64)
constant	19.424	12.715	-1.374	-26.342
	(2.29)	(1.77)	(-0.09)	(-1.49)
R^2	0.298	0.120	0.644	0.486

Notes: The results are OLS estimations. Observations are weighted using sampling weights. Standard errors are adjusted for clustering across strata and sampling units. The t-statistics are in the brackets.

Table 4: Summary of the marginal effects of power within an interactive time allocation model for the median couple

Dependent Variables:		*Housework (minutes)*		*Market work (probabilities)*				
		Wife	*Husband*	*Wife*				*Husband*
Independent Variables				(0 hrs)	(1-39 hrs)	(40-44 hrs)	(>44 hrs)	(>44 hrs)
			Observed frequencies:	0.08	0.26	0.44	0.22	0.55
Housework	*wife*	60	**12 *****	**0.03 *****	**0.04 *****	-0.01	**-0.05 *****	-0.01
	husband	7	60	**-0.01 ****	**-0.02 *****	0.01	**0.02 *****	0.01
Market work	*wife*	**-8 *****	**4 ****	-	-	-	-	**-0.01 ****
	husband	**4 ***	**-5 *****	0.00	0.00	0.00	0.00	-
Age power	*less*	-31	26	0.00	0.00	0.00	0.00	0.06
	more	-68	29	-0.01	-0.02	0.00	0.03	-0.01
Education power	*less*	-52	15	**0.03 ***	**0.04 ***	-0.02	**-0.05 ***	-0.03
	more	47	-13	0.01	0.02	-0.01	-0.02	-0.03
Sex power	*less*	-9	11	0.01	0.01	0.00	-0.01	0.03
	more	-14	**78 ***	-0.02	-0.03	0.01	0.04	**-0.10 ***
Wage power	*less*	**99 ****	**-35 ***	**0.12 *****	**0.10 *****	**-0.09 ****	**-0.13 *****	-0.02
	more	-38	-43	0.00	0.00	0.00	0.00	0.03

Notes: The marginal effects for the housework (market work) variables are calculated using the estimated coefficients of OLS (ordered probit) regressions and are expressed in minutes (probabilities). The median couple is composed of a wife who does 40 hours of market work and 15 hours of housework, and a husband who does 48 hours of market work and 6 hours of housework. */**/*** represent statistical significance at the 90%, 95% and 99% levels, respectively. Statistically significant results are indicated in bold.

Table 5: The mechanisms of intraspousal time allocation, in function of power

	Housework equations		Market work equations	
	Wife	Husband	Wife	Husband
Personal characteristics of the individual				
age	0.260	-0.112	0.003	0.017
	(4.94)	(-4.1)	(0.28)	(1.87)
age2	-0.004	-0.007	0.000	0.001
	(-0.73)	(-2.43)	(-0.38)	(1.53)
edudummy2	-0.293	1.526	-0.172	0.081
	(-0.19)	(2.82)	(-1.27)	(0.53)
edudummy3	0.017	2.502	-0.274	0.180
	(0.01)	(3.37)	(-1.83)	(0.82)
race2	-2.438	0.286	-0.048	-0.346
	(-2.82)	(0.64)	(-0.42)	(-2.24)
race3	-2.417	0.640	-0.268	-0.148
	(-1.38)	(1.07)	(-1.47)	(-0.59)
religion2	-0.006	-0.278	-0.104	0.061
	(-0.01)	(-0.77)	(-0.97)	(0.81)
religion3	-0.884	-1.027	-0.035	-0.060
	(-0.93)	(-1.34)	(-0.2)	(-0.25)
religion4	-3.313	-0.201	-0.551	-0.012
	(-2.67)	(-0.27)	(-2.36)	(-0.07)
religion5	-0.166	-1.204	-0.118	-0.420
	(-0.1)	(-1.77)	(-0.84)	(-1.79)
religion6	-2.167	0.422	-0.131	-0.142
	(-3.03)	(0.71)	(-1.11)	(-1.11)
Labor market characteristics of the individual				
lwagehat	-3.655	0.360	0.271	0.132
	(-2.83)	(0.51)	(1.42)	(0.78)
lwagehat2	-3.828	0.384	-0.084	0.110
	(-0.93)	(0.25)	(-0.24)	(0.47)
*lwagehat*job2*	-1.856	–	–	–
	(-3.95)			
*lwagehat*job3*	-1.987	–	–	–
	(-3.60)			
*lwagehat*job4*	-2.118	-0.342	–	–
	(-3.79)	(-2.60)		
*lwagehat2*job2*	6.810	–	–	–
	(1.40)			
*lwagehat2*job3*	7.161	–	–	–
	(1.19)			
*lwagehat2*job4*	4.078	0.422	–	–
	(0.50)	(0.24)		
weeksdummy2	–	–	0.576	6.427
			(3.23)	(13.59)
weeksdummy3	–	–	0.592	6.532
			(3.5)	(13.8)
unemployment rate	4.520	5.171	0.369	-6.113
	(0.20)	(0.26)	(0.09)	(-1.32)

Continued on the next page

Table 5 (continued)

	Housework equations		Market work equations	
	Wife	Husband	Wife	Husband
Characteristics of the couple				
child1	2.070	-0.243	0.001	0.146
	(2.77)	(-0.54)	(0.01)	(1.56)
child2	3.753	0.137	-0.074	0.174
	(4.50)	(0.24)	(-0.81)	(1.13)
child3	4.107	-0.661	0.132	-0.064
	(3.19)	(-0.96)	(0.9)	(-0.41)
youngchild	1.834	-0.483	-0.226	-0.046
	(2.59)	(-1.00)	(-2.04)	(-0.39)
rooms	-0.161	0.233	–	–
	(-1.13)	(2.57)		
cars	-0.335	0.219	–	–
	(-1.18)	(1.15)		
agepwrlo	-0.518	0.430	-0.001	-0.028
	(-0.75)	(1.09)	(-0.02)	(-0.32)
agepwrhi	-1.127	0.483	0.084	0.158
	(-1.38)	(1.22)	(0.74)	(1.04)
edupwrlo	-0.859	0.248	-0.179	-0.081
	(-1.19)	(0.74)	(-1.98)	(-0.93)
edupwrhi	0.784	-0.219	-0.082	-0.065
	(1.09)	(-0.59)	(-0.95)	(-0.72)
sexratiopwrlo	-0.146	0.183	-0.047	-0.246
	(-0.18)	(0.36)	(-0.46)	(-2.05)
sexratiopwrhi	-0.232	1.300	0.133	0.081
	(-0.35)	(1.98)	(1.28)	(0.66)
wagepwrlo	1.643	-0.577	-0.543	0.064
	(2.26)	(-1.83)	(-5.85)	(0.57)
wagepwrhi	-0.641	-0.723	0.013	-0.051
	(-1.06)	(-1.49)	(0.18)	(-0.53)
marketworkhat	-0.131	-0.082	–	–
	(-3.6)	(-3.73)		
houseworkhat	–	–	-0.087	0.015
			(-5.59)	(0.72)
houseworkhat2	–	–	-0.005	–
			(-3.6)	
marketworkhat, spouse	0.069	0.060	0.005	-0.015
	(1.78)	(2.61)	(1.05)	(-2.08)
houseworkhat, spouse	0.111	0.202	0.067	-0.024
	(0.69)	(3.17)	(4.39)	(-1.17)
constant	22.404	5.264	-0.964	-5.846
	(6.68)	(3.09)	(-1.97)	
constant2	–	–	0.021	–
			(0.04)	
constant3	–	–	1.196	–
			(2.54)	

Notes: The housework results are OLS estimations. The market work results are ordered probit estimations. Observations are weighted using sampling weights. Standard errors are adjusted for clustering across strata and sampling units. The t-statistics are in the brackets.